



UNIVERSITAS INDONESIA

**VIRTUAL SCREENING KANDIDAT OBAT BERBASIS
MACHINE LEARNING PADA DATABASE TANAMAN
HERBAL INDONESIA DENGAN BERBAGAI STRATEGI
FEATURE SELECTION**

TESIS

RAHMAN PUJianto

1406596132

**FAKULTAS ILMU KOMPUTER
PROGRAM MAGISTER ILMU KOMPUTER
DEPOK
JUNI 2017**



UNIVERSITAS INDONESIA

**VIRTUAL SCREENING KANDIDAT OBAT BERBASIS
MACHINE LEARNING PADA DATABASE TANAMAN
HERBAL INDONESIA DENGAN BERBAGAI STRATEGI
FEATURE SELECTION**

TESIS

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Magister Ilmu Komputer**

RAHMAN PUJianto

1406596132

**FAKULTAS ILMU KOMPUTER
PROGRAM MAGISTER ILMU KOMPUTER
DEPOK
JUNI 2017**

ABSTRAK

Nama : Rahman Pujiyanto
Program Studi : Ilmu Komputer
Judul : Virtual Screening Kandidat Obat berbasis Machine Learning
pada Database Tanaman Herbal Indonesia dengan Berbagai
Strategi Feature Selection

Penelitian dalam thesis ini akan melakukan *virtual screening* kandidat obat yang terdiri dari dua tahap. Tahap pertama *screening* menggunakan *ligand-based screening* (LBS) lalu dilanjutkan dengan tahap kedua yaitu *screening* menggunakan *molecular docking*. Penelitian ini akan fokus pada *feature selection* pada tahap pertama. Hasil percobaan menunjukan *feature selection* dapat mempercepat proses training dan prediksi tanpa mengurangi performa pengklasifikasi LBS. *Virtual screening* terhadap Database Tanaman Herbal Indonesia telah mengidentifikasi 10 senyawa yang bereaksi dengan HIV-1 protease.

Kata kunci: Feature selection, Virtual screening, SVM

ABSTRACT

Name : Rahman Pujianto
Program : Computer Science
Title : Virtual Screening Kandidat Obat berbasis Machine Learning
pada Database Tanaman Herbal Indonesia dengan Berbagai
Strategi Feature Selection

The research in this thesis will conduct a virtual screening of candidate drugs consisting of two stages. The first stage of screening using ligand-based screening (LBS) and then followed by the second stage of screening using molecular docking. This study will focus on feature selection in the first stage. Experimental results show feature selection can speed up the training and prediction process without reducing the performance of LBS classifier. Virtual screening of Indonesia Herbal Plant Database has identified 10 compounds that react with HIV-1 protease.

Keyword: Feature selection, Virtual screening, SVM

DAFTAR ISI

ABSTRAK	i
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Sistematika Penulisan	3
2 LANDASAN TEORI	5
2.1 Perancangan Obat	5
2.2 Dinamika Molekuler	6
2.3 Molecular Docking	7
2.4 Klasifikasi	8
2.5 Support Vector Machine	9
2.6 Feature Selection	9
2.6.1 Wrapper Method	10
2.6.2 SVM Recursive Feature Elimination	11
3 METODOLOGI PENELITIAN	12
4 HASIL DAN PEMBAHASAN	16
4.1 Pengumpulan Data dan Preprocessing	16
4.2 Feature Selection	16
4.3 Training Pengklasifikasi dan Virtual Screening	19
4.4 Molecular Docking	21
5 PENUTUP	24
5.1 Kesimpulan	24

5.2	Saran	24
REFERENSI		25
LAMPIRAN		28
Lampiran 1. Daftar Feature Terpilih		28
Lampiran 2. Daftar Senyawa yang Diklasifikasi sebagai Senyawa Obat		32

DAFTAR GAMBAR

2.1	Konsep Dasar Dinamika Molekuler (Wibisono, 2012)	7
2.2	Wrapper Method (Tang, Alelyani, & Liu, 2014)	10
3.1	Alur Penelitian	12
4.1	Grafik akurasi populasi tiap generasi saat feature selection menggunakan wrapper method	17
4.2	Grafik jumlah feature terpilih saat feature selection menggunakan wrapper method	18
4.3	Grafik akurasi berbanding jumlah feature saat feature selection menggunakan SVM-RFE	18
4.4	Grafik ROC untuk SVM-RFE, Wrapper method dan tanpa <i>feature selection</i>	19
4.5	Grafik akurasi, sensitivity, precision, specificity untuk SVM-RFE, Wrapper method dan tanpa <i>feature selection</i>	20
4.6	Protein target dalam docking, yaitu HIV-1 protease	22
4.7	Konformasi senyawa Withangulatin A dengan protein HIV-1 protease	23

DAFTAR TABEL

3.1	Jenis <i>molecular description</i> yang dapat diextract oleh <i>PaDEL Description</i>	13
4.1	Perbandingan waktu yang diperlukan untuk learning dan prediksi masing-masing pengklasifikasi dalam detik	20
4.2	Perbandingan dengan penelitian (Yanuar, Suhartanto, Mun'im, Anugrah, & Syahdi, 2014)	21
4.3	Sepuluh senyawa dengan binding energy terendah beserta species asal senyawa tersebut.	22

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Berkembangnya penyakit baru dan penyakit lama yang menjadi resistant terhadap obat yang ada menimbulkan kebutuhan akan obat-obat baru. Sebagai contoh HIV/AIDS yang menjadi masalah kesehatan global. HIV memiliki resistansi obat yang tinggi karena memiliki tingkat replikasi dan mutasi yang tinggi. Tetapi pembuatan obat merupakan proses yang rumit dan lama, sehingga sering kali tidak dapat memenuhi kebutuhan obat yang ada.

Kandidat obat merupakan sebagian kecil dari keseluruhan senyawa yang ada. Untuk mencari kandidat-kandidat ini proses pencarian obat seperti *high throughput screening (HTS)* memeriksa secara kimiawi reaksi berbagai senyawa pada target obat. Tetapi proses ini memiliki efisiensi rendah dan mahal karena sebagian besar senyawa yang habis terpakai dalam percobaan. Untuk itu perlu ada upaya untuk membatasi jumlah senyawa yang akan diperiksa (Chen & Visco Jr., 2016).

Virtual screening menggunakan algoritma komputer dalam mencari senyawa kimia yang memiliki kemungkinan bereaksi terhadap target obat. Sehingga *virtual screening* dapat digunakan untuk mengurangi jumlah senyawa yang akan dites menggunakan percobaan kimia, yang pada akhirnya mengurangi waktu dan biaya yang dibutuhkan untuk mencari kandidat obat (Korkmaz, Zararsiz, & Goksuluk, 2014).

Virtual screening memiliki berbagai teknik yang dikembangkan, antara lain simulasi dinamika molekuler, *molekuler docking*, dan *ligand-based screening (LBS)*. Simulasi dinamika molekuler melakukan simulasi reaksi kimia di komputer. *Molekuler docking* bekerja dengan cara menentukan apakah suatu senyawa dapat terikat pada suatu protein target obat. Baik simulasi dinamika molekuler maupun molecular docking membutuhkan kekuatan komputasi yang besar. LBS bekerja dengan cara mencari kesamaan struktur kimia maupun

karakteristik kimia suatu senyawa dengan senyawa obat yang telah ada. Agar dapat bekerja LBS membutuhkan informasi mengenai senyawa obat yang telah ada. Informasi tersebut bisa didapatkan dari hasil HTS yang tersedia melalui berbagai database seperti PubChem BioAssay (Wang et al., 2014), ChEMBL (Bento et al., 2014), PubChem Compound (Kim et al., 2016) dan ZINC (Irwin, Sterling, Mysinger, Bolstad, & Coleman, 2012).

Penelitian terkait oleh Hilman, 2012 secara garis besar melakukan *virtual screening* dalam dua tahap. Tahap pertama LBS memilih senyawa-senyawa potensial dari keseluruhan senyawa yang hendak di-*screening*. Penelitian Hilman, 2012 mendapati klasifikasi menggunakan SVM merupakan metode yang unggul untuk digunakan sebagai LBS. Pada tahap kedua *molekuler docking* akan melakukan screening pada senyawa-senyawa potensial untuk mendapatkan kandidat obat. Pemilihan senyawa-senyawa potensial oleh LBS pada tahap pertama perlu dilakukan mengingat tingginya sumber daya komputasi yang diperlukan oleh *molekuler docking*. Tahap pertama merupakan tahap yang penting, karena senyawa-senyawa potensial dapat luput apabila tahapan ini tidak berjalan dengan baik (Korkmaz et al., 2014).

Molecular descriptor merupakan perhitungan yang mengubah informasi yang terkandung dalam simbol struktur senyawa kimia menjadi bilangan yang berguna, *molecular descriptor* juga dapat berasal dari percobaan yang telah baku (Yap, 2010). LBS menggunakan *molecular descriptor* sebagai *feature* dalam klasifikasi. *Molecular descriptor* memegang peran penting dalam LBS. Dari suatu senyawa kimia dapat diekstrak setidaknya 701 jenis *molecular descriptor* (Yap, 2010). Jumlah yang banyak ini memperumit dalam memahami hubungan antara *molecular descriptor* suatu senyawa dengan potensi senyawa tersebut menjadi obat (Korkmaz et al., 2014). Pengamatan awal pada dataset yang hendak digunakan terlihat tidak semua *molekuler descriptor* memberikan kontribusi terhadap hasil klasifikasi. Sebagai contoh jumlah atom Bromin (Br), yang merupakan salah satu *molecular descriptor*, adalah 0 untuk semua senyawa dalam dataset dan ada sekitar 500 *molecular descriptor* yang seperti ini.

Dalam penelitian ini penulis akan meningkatkan akurasi LBS. Salah satu cara untuk meningkatkan akurasi adalah menggunakan *Feature Selection* (Korkmaz et al., 2014). *Feature selection* merupakan cara untuk mengurangi *Feature selection* bermanfaat untuk meningkatkan akurasi, menghasilkan prediktor yang lebih cepat dari segi *computational cost*, serta memberikan pemahaman yang lebih baik terhadap model yang dihasilkan (Janecek, Gansterer, Demel,

& Ecker, 2008). Hal tersebut membuat *feature selection* menjadi langkah yang penting dalam berbagai metode *machine learning*. Akan tetapi peningkatan akurasi yang dicapai oleh berbagai strategi *feature selection* sangat bergantung pada jenis data (Janecek et al., 2008).

Dalam penelitian ini, untuk *training set* akan digunakan PubChem Bioassay, suatu basis data yang disediakan oleh NCBI yang menyediakan deskripsi *bioassay* termasuk kondisi dan prosedur *screening* dari beragam senyawa. Selanjutnya dalam penelitian ini akan melakukan *virtual screening* terhadap dataset senyawa zat alami Tanaman Obat Herbal Indonesia. Database Tanaman Obat Herbal Indonesia merupakan hasil dari upaya Departemen Farmasi FMIPA UI untuk mengumpulkan struktur 3D senyawa-senyawa kimia yang terkandung dalam Tanaman Herbal Indonesia.

1.2 Perumusan Masalah

Berdasarkan penjabaran diatas dapat disimpulkan pertanyaan riset yang perlu dijawab adalah, strategi *feature selection* apa yang dapat meningkatkan akurasi LBS kandidat obat pada database tanaman herbal indonesia? Bagaimana hasilnya apabila dibandingkan dengan *molekular docking*?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk mencari strategi *feature selection* yang dapat digunakan untuk meningkatkan akurasi LBS.

1.4 Sistematika Penulisan

Penelitian ini disusun dengan sistematika penulisan sebagai berikut.

- **Bab 1 Pendahuluan**

Pada bab ini akan dibahas latar belakang dari topik penelitian yang dilakukan, merumuskan masalah yang menjadi kata kunci dari penelitian dan membahas tujuan. Selain itu, sistematika dalam penulisan tesis menjadi pembahasan terakhir yang menutup bab pendahuluan ini

- **Bab 2 Landasan Teori**

Pada bab ini akan dibahas teori dan teknologi perancangan obat. Pada bagian ini berisi tentang sesuatu masalah yang diteliti yang kebenarannya telah terbukti melalui data yang diperoleh lewat penggunaan metode penelitian yang relevan dan terpercaya. Selain itu pada bagian ini dijelaskan mengenai skema ringkas yang akan dijadikan acuan dalam pengumpulan data di lapangan.

- **Bab 3 Metodologi Penelitian**

Pada bab ini akan dibahas tahapan dari penelitian yang dilakukan. Selain itu, ulasan mengenai data dan seluruh kegiatan yang akan dilakukan pada penelitian. Bahasan yang tidak kalah penting dari bagian ini adalah *tools* dan aplikasi perangkat lunak yang mendukung.

- **Bab 4 Hasil dan Pembahasan**

Bab ini akan menjelaskan tentang hasil eksperimen dan analisis hasil eksperimen.

- **Bab 5 Penutup**

Bab ini akan menjelaskan tentang kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian berikutnya.

BAB 2

LANDASAN TEORI

2.1 Perancangan Obat

Proses pembuatan merupakan proses yang lama dan sulit, sering kali tidak dapat mengimbangi kebutuhan yang ada, harus melalui berbagai uji coba obat yang sangat ketat untuk menyatakan keamanan dan keefektifannya sebelum dapat tersedia secara umum. Bahkan sebelum uji coba obat, perlu dicari kandidat obat dan banyak tantangan untuk menemukannya. Senyawa kimia yang dapat dijadikan kandidat obat sangat banyak, sementara untuk suatu target obat hanya sebagian kecil saja yang dapat menjadi kandidatnya. (Chen & Visco Jr., 2016)

High throughput screening (HTS) merupakan metode pertama yang secara sistematis memeriksa dan mengidentifikasi kandidat obat dari banyak senyawa kimia dalam waktu singkat. HTS lebih efektif dari metode sebelumnya, memeriksa lebih banyak senyawa kimia dari yang pernah ada, tetapi HTS sangat mahal untuk dilakukan, karena senyawa kimia yang diperiksa akan habis terpakai pada prosesnya. (Chen & Visco Jr., 2016)

Walaupun demikian HTS menghasilkan banyak data. Tidak hanya lebih banyak data yang dihasilkan, tetapi data-data ini juga lebih banyak dibagikan. Keterseediaan data ini menghasilkan cara-cara baru untuk mencari kandidat obat. Sekarang memungkinkan untuk melakukan HTS secara virtual untuk mencari senyawa yang memiliki kemungkinan besar untuk menjadi kandidat obat. Cara ini meningkatkan efisiensi karena hasil sebelumnya digunakan sebagai panduan dalam mencari hasil yang baru, meningkatkan laju penemuan kandidat baru dan mengurangi sumber daya yang diperlukan. (Chen & Visco Jr., 2016)

Virtual screening adalah sebuah proses pencarian kandidat obat dengan menganalisa senyawa kimia dengan bantuan komputer. *Virtual screening* memiliki berbagai teknik yang dikembangkan, antara lain simulasi dinamika molekuler, *molekuler docking*, dan *ligand-based screening* (LBS). (Chen & Visco Jr.,

2016)

Simulasi dinamika molekuler dan *docking* merupakan proses komputasi numerik dalam pencarian model geometri dan energi yang cocok pada pasangan *ligand* dan target obat. Proses ini memerlukan tenaga komputasi yang besar dan terkadang tergantung parameter awal serta perlu dijalankan beberapa kali. Kelebihan cara ini adalah tidak diperlukannya informasi dari percobaan sebelumnya. (Chen & Visco Jr., 2016)

LBS menggunakan informasi ikatan yang telah ada antara senyawa kimia kandidat obat atau sering disebut *ligand* dengan target obat untuk mengidentifikasi *ligand* lain yang mungkin dapat mengikat target obat. LBS secara sederhana bekerja dengan mencari senyawa yang memiliki struktur yang sama dengan *ligand* yang telah diketahui atau pun dengan klasifikasi yang lebih rumit berdasarkan ciri-ciri kimia dan strukturnya. LBS tidak membutuhkan tenaga komputasi sebesar simulasi molekuler, tetapi membutuhkan informasi mengenai pasangan *ligand* dan target obat yang telah ada. (Chen & Visco Jr., 2016)

Gabungan antara keduanya juga dapat digunakan, seperti yang telah dilakukan oleh Hilman, 2012. Dimana LBS digunakan sebagai langkah awal untuk mengurangi jumlah senyawa yang akan *discreening* menggunakan *molecular docking*. Walaupun cara ini tetap memerlukan informasi mengenai pasangan *ligand* dan target obat yang telah ada, tetapi cara ini dapat mempersingkat waktu yang dibutuhkan untuk *virtual screening*. Cara ini sangat bergantung pada proses *pre-screening*, karena *pre-screening* yang buruk mengakibatkan luputnya senyawa-senyawa yang potensial. (Korkmaz et al., 2014)

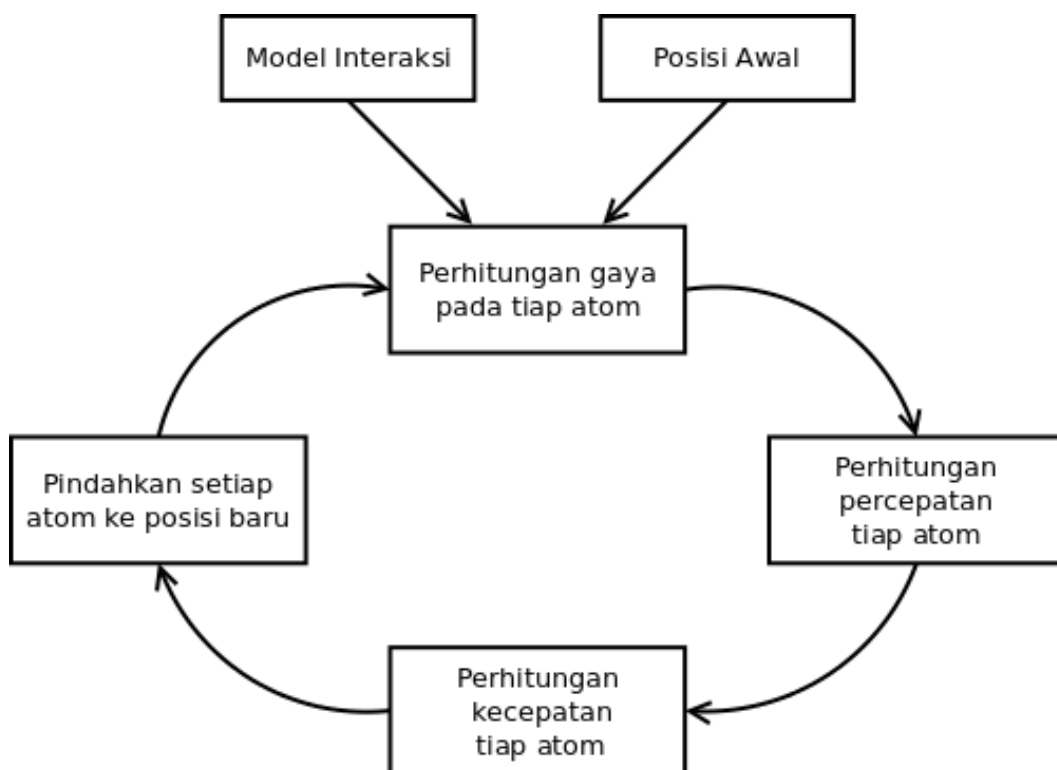
2.2 Dinamika Molekuler

Metode dinamika molekuler diperkenalkan pertama kali oleh Alder dan Wainwright pada tahun 1950 yang digunakan untuk melakukan simulasi interaksi pada *hard spheres*. Pada tahun 1964 Rahman, A. melakukan simulasi cairan argon. Simulasi protein pertama dilakukan pada tahun 1977 oleh McCammon. Saat ini telah banyak digunakan dalam studi mengenai pelarutan protein, mempelajari ikatan *ligand*, interaksi kompleks DNA-protein, dan pelipatan protein. (Wibisono, 2012)

Dinamika molekuler berfungsi untuk menunjukkan struktur molekul, pergerakan dan fungsi molekul. Dinamika molekuler menggunakan mekanika molekul untuk

melakukan komputasi pergerakan setiap atom. Pergerakan ini merupakan reaksi yang dihasilkan oleh struktur molekul dan merupakan element penting dari fungsi khusus yang dimiliki oleh molekul. Dinamika molekuler dapat membantu dalam memahami bagaimana suatu molekul bekerja. (Suhartanto, Yanuar, Hilman, Wibisono, & Dermawan, 2012)

Dinamika molekuler bekerja pertama-tama dengan menentukan gaya total dari setiap atom yang disimulasikan berdasarkan interaksinya dengan atom-atom lain. Dengan gaya tersebut dapat dihitung percepatan setiap atomnya, yang dapat digunakan untuk mendapatkan kecepatan. Proses selanjutnya adalah memindahkan atom keposisi barunya lalu kembali ke langkah awal. (Wibisono, 2012)



Gambar 2.1: Konsep Dasar Dinamika Molekuler (Wibisono, 2012)

2.3 Molecular Docking

Ligand-protein docking adalah proses dimana lingand dan protein membentuk suatu *complex*. Tugas *molecular docking* adalah untuk memprediksi ikatan yang benar antara *ligand* dengan protein, yaitu dengan mencari *complex* geometri antara *ligand* dengan protein yang memiliki energi bebas terendah (Sotriffer

et al., 2000). Molecular docking digunakan pada proses pencarian obat secara virtual screening untuk mencari ligand terbaik yang dapat terikat pada protein target dari penyakit (Swardiana, 2014). *Molecular docking* terdiri dari dua fungsi utama, pertama *search function* yang digunakan untuk mengeksplorasi geometri-geometri ikatan yang mungkin terjadi. Kedua *scoring function* untuk mengevaluasi geometri ikatan tersebut (Huey, Morris, Olson, & Goodsell, 2007).

AutoDock adalah software untuk melakukan *docking* secara otomatis molekul kecil seperti *peptides*, *enzyme inhibitor*, dan obat, ke *macromolecules*, seperti *protein*, *enzyme*, *antibody*, DNA, dan RNA (Sotriffer et al., 2000). Autodock memanfaatkan *Lamarckian genetic algorithms* sebagai *search function* dan *semi-empirical free energy force field* sebagai *scoring function* (Morris et al., 1998) (Huey et al., 2007).

2.4 Klasifikasi

Tujuan dari klasifikasi adalah untuk menentukan kelas suatu pola yang belum diketahui kelasnya (Theodoridis, 2015). Sebagai contoh adalah menentukan apakah suatu senyawa termasuk kedalam kelompok yang bereaksi terhadap target obat atau kelompok yang tidak bereaksi. Langkah pertama dalam melakukan klasifikasi adalah menentukan bagaimana cara merepresentasikan pola, dan umum dilakukan pada tahap *preprocessing*. Pola harus dapat di-representasikan secara efisien dan kaya informasi. Hal ini dilakukan dengan mengubah data mentah masing-masing pola kedalam vector $\mathbf{x} \in \mathbb{R}^l$. Vektor ini disebut *feature vector* dan setiap elemen l merupakan *feature*. Pada awalnya terdapat banyak *feature* yang pada akhirnya dipilih yang paling informatif melalui *feature selection* (Theodoridis, 2015)

Setelah menentukan representasi data maka tahap selanjutnya adalah melatih pengklasifikasi. Untuk itu diperlukan data set yang telah diketahui kelasnya untuk dijadikan *training set*. Data set ini berbentuk pasangan, $(y_n, \mathbf{x}_n), n = 1, 2, \dots, N$, dimana y_n menyatakan kelas dari \mathbf{x}_n , dan biasa disebut *class label*. Untuk klasifikasi dua kelas, $y_n \in -1, +1$. Berdasarkan *training data* akan dibentuk fungsi f yang akan digunakan untuk menentukan label dari setiap masukan *feature vector*. Fungsi ini disebut pengklasifikasi. (Theodoridis, 2015)

Setelah pengklasifikasi didapatkan maka siap untuk melakukan klasifikasi. *Feature vector* dari pola yang tidak diketahui kelasnya akan dijadikan masukan untuk fungsi pengklasifikasi, hasil outputnya akan berupa prediksi kelas untuk pola yang bersangkutan.

2.5 Support Vector Machine

Support Vector Machine (SVM) merupakan metode klasifikasi yang populer, pertama dikenalkan oleh Vapnik beserta koleganya dan telah menarik perhatian berbagai komunitas sains karena latar belakang matematika yang kuat dan kesuksesannya. SVM juga mampu melakukan klasifikasi nonlinear dan mampu mengatasi dimensi data yang tinggi, sehingga digunakan pada banyak bidang, seperti bioinformatika, klasifikasi text, pengolahan citra, dan klasifikasi kanker. (Korkmaz et al., 2014)

Pada problem klasifikasi binary, misal $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ merupakan training data dengan vector n-dimensi dalam ruang ($\mathbf{x}_i \in \mathbb{R}^n$) dan $\{y_1, \dots, y_n\}$ merupakan labelnya, dimana $y_i \in \{-1, +1\}$. Tujuan disini adalah untuk mencari sebuah *hyperplane* dan mendapatkan persamaan yang dapat memisahkan data *training* menjadi dua bagian, dimana semua data dengan kelas yang sama menempati sisi *hyperplane* yang sama. Data yang terdekat dengan *hyperplane* pada kedua sisinya disebut dengan *support vectors* dan tujuan dari SVM adalah untuk memaksimalkan jarak antara *support vector* dengan *hyperplane*.

Bila data tidak dapat dipisahkan secara linear, maka *slack variable* dapat digunakan untuk membuat beberapa data menjadi *misclassified*. Lebih jauh lagi apabila data dapat dipisahkan secara nonlinear, SVM mengatasinya dengan memetakan vektor input ke ruang dimensi yang lebih besar dengan memanfaatkan *kernel function*.

2.6 Feature Selection

Dengan meningkatnya dimensi data, berbagai jenis analisa data dan problem klasifikasi menjadi semakin sulit. Terkadang data tersebut menjadi sangat menyebar di ruang yang ditempatinya. Hal ini dapat menjadi masalah bagi *supervised learning* maupun *unsupervised learning*. Fenomena ini disebut dengan *curse of dimensionality*. (Janecek et al., 2008)

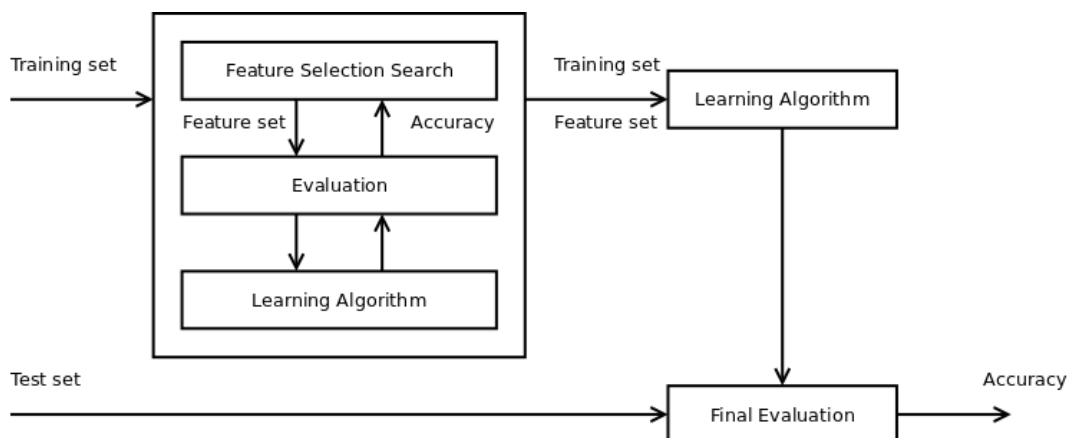
Pada problem klasifikasi, *overfitting* terjadi pada saat fitur berjumlah banyak dan data yang digunakan untuk *training* relatif sedikit. Pada situasi seperti itu mudah untuk mendapatkan model yang memiliki performa yang baik pada training data tetapi buruk saat digunakan pada data testing.

Feature selection bertujuan untuk mencari subset dari fitur-fitur dataset. Dimana dengan menggunakan subset fitur ini dapat dihasilkan model klasifikasi dengan akurasi tertinggi. Berbeda dengan *feature extraction*, *feature selection* memilih subset dari fitur yang sudah ada, bukan membentuk fitur baru.

Penelitian ini akan menggunakan dua teknik *feature selection*, yaitu *Wrapper Metode* dan *SVM Recursive Feature Elimination* (SVM-RFE). Tidak seperti *feature selection* menggunakan *filter method* dimana *feature* dievaluasi secara terpisah, SVM-RFE dan *Wrapper Method* mengevaluasi langsung sekelompok *feature-feature* sehingga Penulis meyakini dua metode tersebut dapat memilih *feature set* yang lebih optimal.

2.6.1 Wrapper Method

Pada *wrapper method*, *feature selection* dilakukan memanfaatkan algoritma *machine learning* sebagai *black box*, dengan kata lain pengetahuan bagaimana cara algoritma tersebut bekerja tidak diperlukan, hanya antarmukanya saja. *Wrapper method* memanfaatkan *machine learning* untuk mengevaluasi kualitas dari fitur-fitur terpilih.



Gambar 2.2: Wrapper Method (Tang, Alelyani, & Liu, 2014)

Bagian *feature selection search* akan menghasilkan kumpulan fitur-fitur, dan bagian *feature evaluation* akan memanfaatkan *classifier* yang dihasilkan oleh

machine learning untuk mengukur performanya, yang selanjutnya digunakan oleh *feature selection search* pada iterasi pencarian kumpulan fitur-fitur selanjutnya. Kumpulan fitur-fitur dengan performa tertinggi akan digunakan untuk menghasilkan *classifier* akhir. *Classifier* ini akan dievaluasi menggunakan data set testing yang tidak digunakan pada proses *learning*. (Tang, Alelyani, & Liu, 2014)

2.6.2 SVM Recursive Feature Elimination

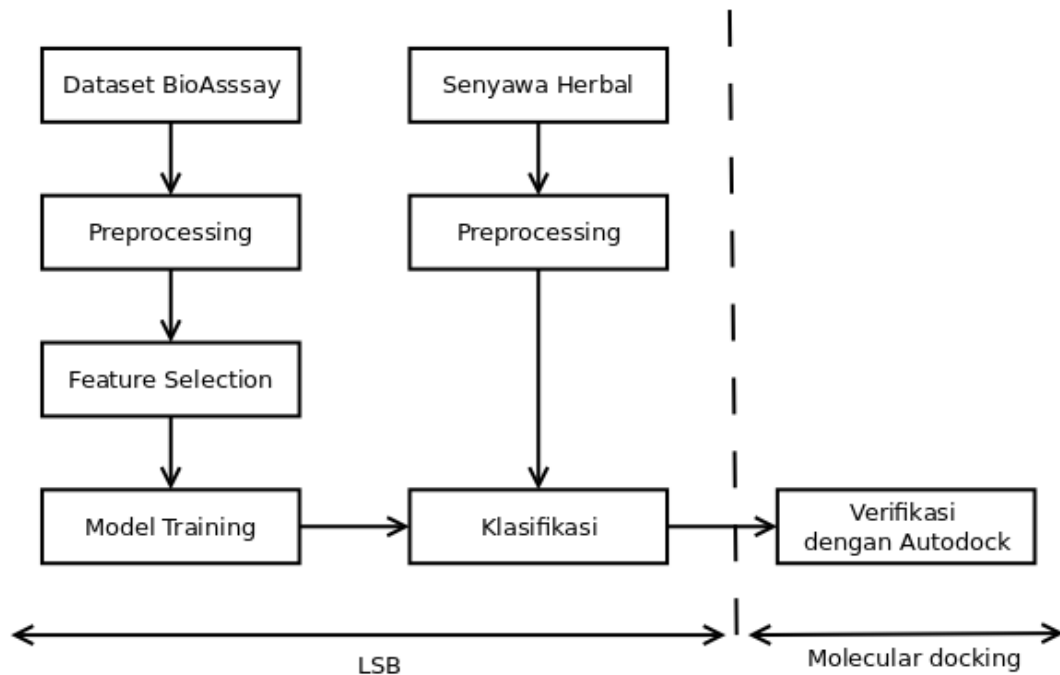
SVM Recursive Feature Elimination (SVM-RFE) pertama kali diajukan oleh Guyon, Weston, and Barnhill, 2002 untuk memilih gen yang relevan pada problem klasifikasi kanker. SVM-RFE memanfaatkan bobot pada *hyperplane* yang dihasilkan sebagai ranking fitur. SVM-RFE terdiri dari empat langkah berikut:

1. Melatih SVM menggunakan data set *training*.
2. Urutkan fitur berdasarkan bobot pada pengklasifikasi yang dihasilkan.
3. Buang fitur dengan bobot terkecil.
4. Ulangi proses diatas menggunakan fitur-fitur yang tersisa.

BAB 3

METODOLOGI PENELITIAN

Pada penelitian ini penulis melakukan *virtual screening* yang terdiri dari dua tahapan, tahap pertama dilakukan *screening* menggunakan LBS lalu untuk senyawa-senyawa yang lolos dilanjutkan dengan *screening* tahap kedua menggunakan *molecular docking*. Secara garis besar alur dari penelitian ini dapat digambarkan sebagai berikut.



Gambar 3.1: Alur Penelitian

Penelitian ini akan menggunakan dua dataset, yaitu *dataset* yang berasal dari PubChem Bioassay dan *dataset* senyawa zat alami tanaman obat herbal Indonesia.

Dataset PubChem BioAssay akan digunakan untuk melatih pengklasifikasi. Akan digunakan BioAssay yang memiliki target *HIV-1 protease* (GI:75593047), seperti AID 162030, AID 160444 dan AID 83109. BioAssay tersebut merupakan hasil percobaan untuk mencari senyawa *HIV-1 protease inhibitor*. Hal ini dilakukan agar sejalan dengan penelitian (Yanuar, Suhartanto, Mun'im, Anugrah,

& Syahdi, 2014) untuk mencari *HIV-1 protease inhibitor* pada tanaman obat herbal indonesia.

Dataset kedua berasal dari Database Tanaman Herbal Indonesia (Yanuar et al., 2011). Database ini merupakan upaya Departemen Farmasi FMIPA UI untuk mengumpulkan struktur 3D senyawa-senyawa kimia yang terkandung dalam tanaman herbal. Sampai saat ini pada database tersebut telah terkumpul tidak kurang dari 1412 struktur senyawa kimia dari tanaman herbal Indonesia. Akan dicari senyawa yang dapat menghambat (*inhibitor*) *HIV-1 protease* dari senyawa-senyawa yang terdapat pada dataset ini.

Langkah pertama yang dilakukan adalah preprocessing untuk mendapatkan *feature* dari setiap senyawa. Pada langkah ini *molecular description* setiap senyawa akan diekstrak untuk digunakan sebagai *feature*. Ekstraksi ini dilakukan menggunakan *PaDEL Description* (Yap, 2010). Sebanyak 667 *molecular description* akan diekstrak oleh *PaDEL Description*, tabel 3.1 menunjukkan jenis *molecular description* yang diekstrak.

Tabel 3.1: Jenis *molecular description* yang dapat diekstrak oleh *PaDEL Description*.

Descriptor type	Number	Class
ALOGP	3	2D
APol	1	2D
Aromatic atoms counts	1	2D
Aromatic bonds count	1	2D
Atom count	13	2D
Autocorrelation (charge)	5	2D
Autocorrelation (mass)	5	2D
Autocorrelation (polarizability)	5	2D
BCUT	6	2D
Bound count	5	2D
BPol	1	2D
Carbon types	9	2D
Chi chain	10	2D
Chi cluster	8	2D
Chi path	16	2D
Chi path cluster	6	2D
Eccentric connectivity index	1	2D

Descriptor type	Number	Class
Atom type electrotopological state	482	2D
Fragment complexity	1	2D
Hbond acceptor count	1	2D
Hbond donor count	1	2D
Kappa shape indices	3	2D
Largest chain	1	2D
Largest Pi system	1	2D
Longest aliphatic chain	1	2D
Mannhold LogP	1	2D
McGowan volume	1	2D
Molecular distance edge	19	2D
Molecular linear free energy relation	6	2D
Petitjean number	1	2D
Ring count	34	2D
Rotatable bonds count	1	2D
Rule of five	1	2D
Topological polar surface area	1	2D
Vertex adjacency information (magnitude)	1	2D
Weight	1	2D
Weighted path	5	2D
Wiener numbers	2	2D
XlogP	1	2D
Zagreb index	1	2D

Setelah *feature* setiap senyawa didapatkan maka selanjutnya dilakukan *feature selection*. Sebelum itu dataset PubChem BioAssay dibagi menjadi dua bagian, satu bagian digunakan sebagai dataset *training* dan bagian kedua digunakan sebagai dataset validasi. *Feature selection* dilakukan menggunakan tiga strategi yaitu:

1. Tanpa menggunakan feature selection.
2. Menggunakan wrapper method sebagai feature selection.
3. Menggunakan SVM-RFE sebagai feature selection.

Langkah selanjutnya adalah melakukan *training* model pengklasifikasi untuk setiap strategi *feature selection*. Pada penelitian ini akan menggunakan SVM

untuk *training* model pengklasifikasi. Setiap pengklasifikasi yang dihasilkan akan dievaluasi peformanya menggunakan dataset validasi. Evaluasi peforma dilakukan dengan pengukuran berikut:

$$\begin{aligned}\text{Accuracy rate} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{Sensitivity} &= \frac{TP}{TP + FN} \\ \text{Specificity} &= \frac{TN}{TN + FP} \\ \text{Precision} &= \frac{TP}{TP + FP}\end{aligned}$$

Dimana, TP = true positive, TN = true negative, FP = false positive, dan FN = *false negative*.

Pengklasifikasi dengan peforma terbaik akan digunakan untuk melakukan *screening* senyawa pada dataset tanaman obat herbal Indonesia. Senyawa yang diklasifikasikan dapat menjadi senyawa obat akan *discreening* lebih lanjut menggunakan Autodock. Terakhir senyawa yang lolos *screening* oleh Autodock dapat dibandingkan dengan (Yanuar et al., 2014).

BAB 4

HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data dan Preprocessing

Langkah pertama dari keseluruhan penelitian ini adalah mengumpulkan data yang akan digunakan untuk *training* pengklasifikasi. Pada Pubchem Bioassay dilakukan pencarian senyawa-senyawa yang bereaksi terhadap *HIV-1 protease*. Pencarian ini menghasilkan 3665 senyawa yang akan digunakan dalam proses *training* sebagai kelas obat. Decoy DUD-E (Mysinger, Carchia, Irwin, & Shoichet, 2012) yang juga untuk target *HIV-1 protease* digunakan sebagai kelas bukan obat. Decoy ini berjumlah 36278 senyawa.

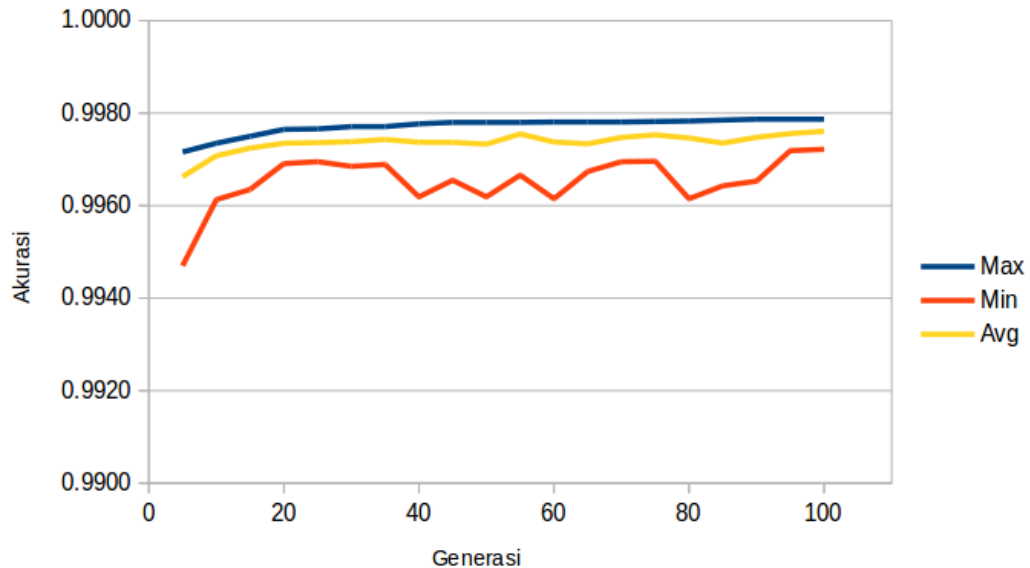
Langkah selanjutnya adalah mengekstrak *molecular description* yang akan digunakan sebagai feature. Untuk mengekstraknya digunakan *PaDEL Description*. Pada saat proses ekstraksi *Molecular description* yang dipilih adalah *molecular description 1D & 2D* lalu *option* yang dipilih adalah untuk menghilangkan garam dan menambah atom hidrogen. Langkah ini menghasilkan 667 *molecular description* yang akan digunakan sebagai *feature*.

Setelah melakukan ekstraksi *feature* ada dua proses yang dilakukan terhadap data-data tersebut. Pertama mengisi *missing value* dengan nol. Kedua, untuk membantu mempercepat proses *feature selection* maka fitur-fitur dengan *variance* nol dihilangkan. *Feature* yang tersisa setelah proses ini berjumlah 486.

4.2 Feature Selection

Selanjutnya dilakukan *feature selection* menggunakan dua metode berbeda, yaitu *wrapper method* dan SVM-RFE. Weka (Hall et al., 2009) digunakan untuk melakukan *wrapper method feature selection*. Sebagai *feature selection search* digunakan *genetic algorithm* dengan ukuran populasi sebesar 20 dan generasi sebanyak 100.

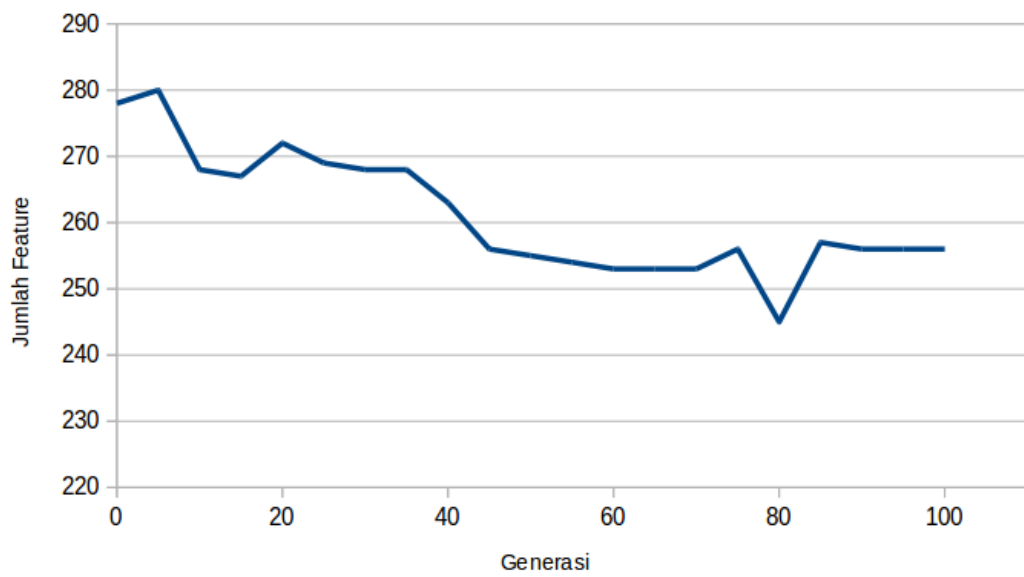
Gambar 4.1 menunjukkan akurasi maksimum, akurasi rata-rata, dan akurasi minimum populasi pada setiap generasi. Terlihat bahwa akurasi sudah sulit untuk meningkat lagi dan ada pada titik maksimumnya yaitu pada nilai 99.78%.



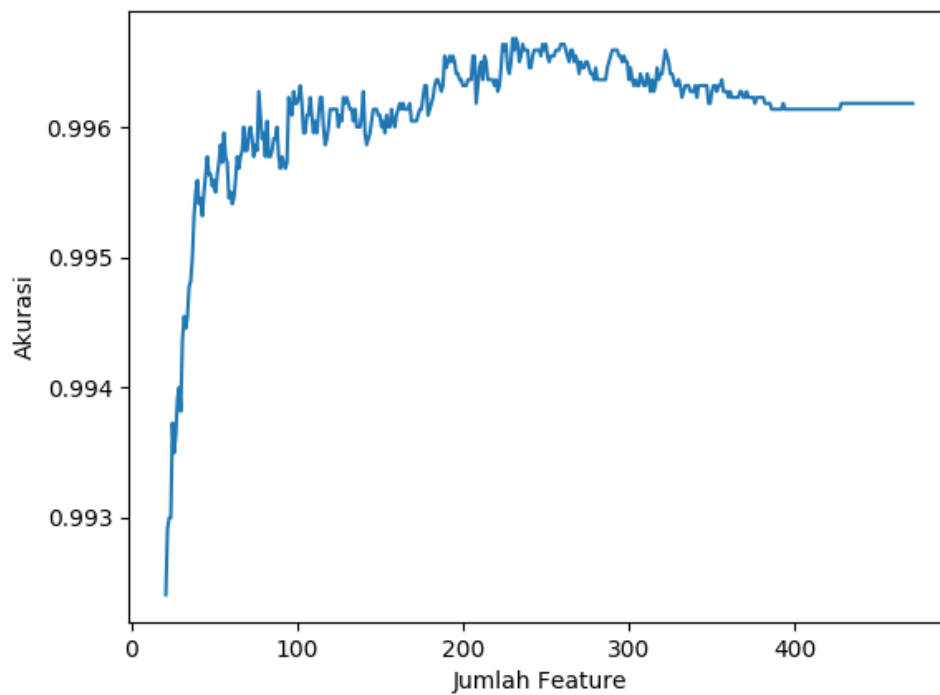
Gambar 4.1: Grafik akurasi populasi tiap generasi saat feature selection menggunakan wrapper method

Gambar 4.2 menunjukkan jumlah *feature* populasi dengan akurasi tertinggi pada setiap generasi. Apabila disejajarkan dengan gambar 4.1, terlihat jumlah *feature* menurun seiring dengan naiknya akurasi. Jumlah *feature* paling optimum berdasarkan wrapper method adalah 256 buah.

Tool scikit-learn (Pedregosa et al., 2011) digunakan untuk melakukan SVM-RFE. SVM-RFE bekerja dengan menggunakan bobot pada *hyperplane* sebagai ranking *feature*. Pada penelitian ini disetiap iterasi SVM-RFE akan dibuang satu *feature* dengan bobot terkecil. Gambar 4.3 menunjukkan akurasi berbanding dengan jumlah *feature* saat menggunakan SVM-RFE. Terlihat akurasi meningkat seiring berkurangnya jumlah *feature*, dan puncaknya pada 232 *feature* dengan akurasi 99.66%.



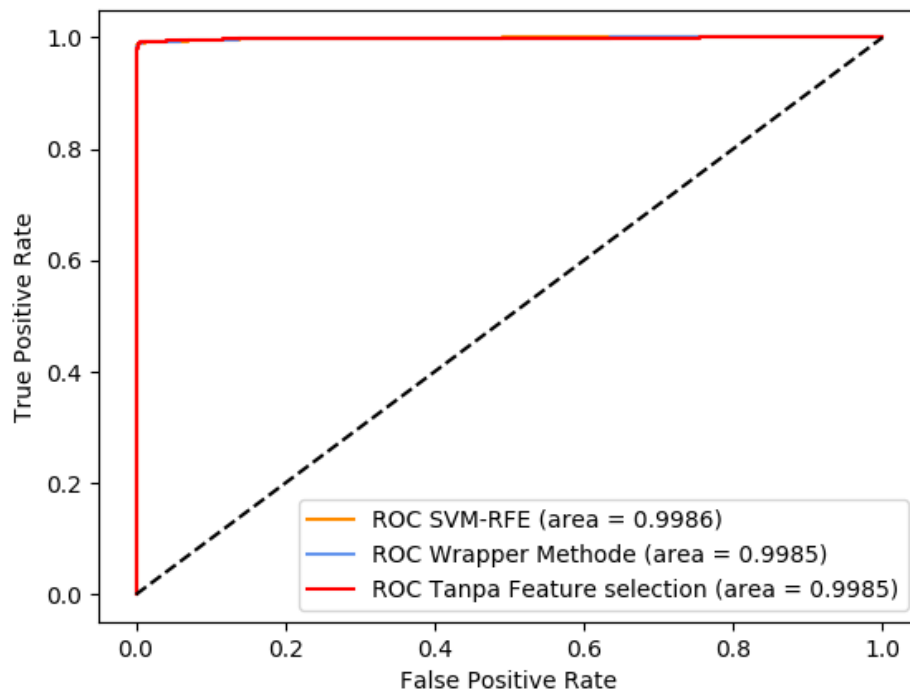
Gambar 4.2: Grafik jumlah feature terpilih saat feature selection menggunakan wrapper method



Gambar 4.3: Grafik akurasi berbanding jumlah feature saat feature selection menggunakan SVM-RFE

4.3 Training Pengklasifikasi dan Virtual Screening

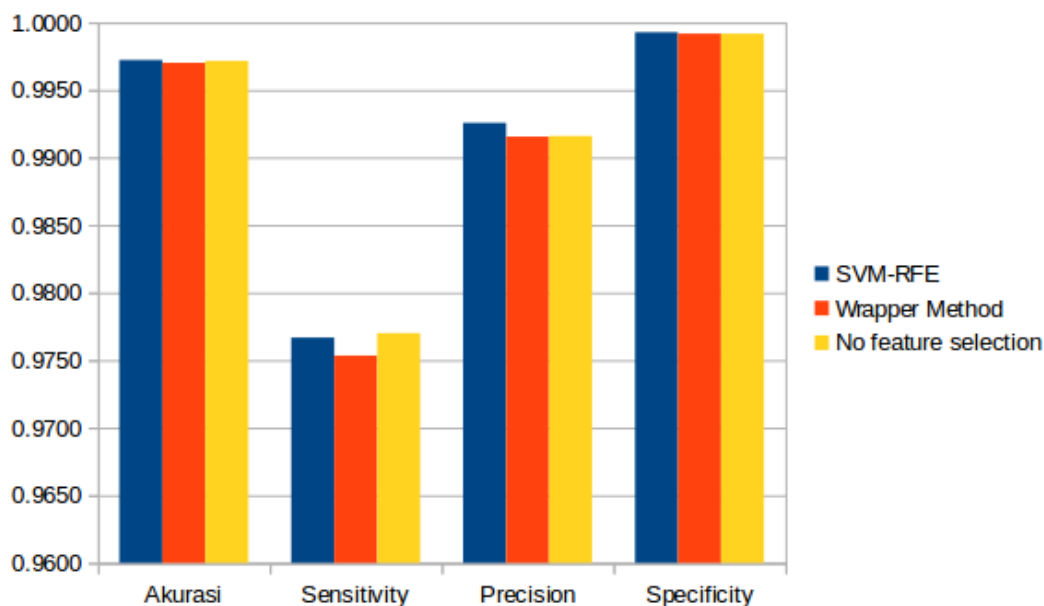
Fitur-fitur yang terpilih oleh SVM-RFE dan fitur-fitur yang terpilih oleh *wrapper method* selanjutnya masing-masing akan digunakan untuk *mentrain* pengklasifikasi menggunakan SVM. Selanjutnya dilakukan validasi terdapat pengklasifikasi-pengklasifikasi tersebut menggunakan *k-fold validation* dengan jumlah *fold* sebanyak 3. Gambar 4.4 menunjukkan grafik ROC untuk masing-masing pengklasifikasi. Terlihat performa pengklasifikasi dengan *feature selection* tidak berkurang dibandingkan dengan pengklasifikasi yang menggunakan seluruh *feature*.



Gambar 4.4: Grafik ROC untuk SVM-RFE, Wrapper method dan tanpa *feature selection*

Gambar 4.5 menunjukkan akurasi, sensitivity, precision, dan specificity setiap pengklasifikasi. Akurasi ketiga pengklasifikasi tidak terpaut jauh dimana SVM-RFE dan tanpa *feature selection* dengan 99.72% sedangkan *wrapper method* memiliki akurasi 99.70%. Sensitivity tertinggi terdapat pada pengklasifikasi tanpa *feature selection* sebesar 97.70% diikuti oleh SVM-RFE sebesar 97.67% lalu *wrapper method* sebesar 97.53%. Sensitivity memiliki nilai yang menarik, karena paling rendah diantara empat indikator lainnya. Sensitivity yang rendah

menunjukkan *false negative* yang lebih banyak, hal ini mungkin disebabkan oleh jumlah data setiap kelas yang tidak *balance*. *unbalance data* dapat mengakibatkan pengklasifikasi lebih condong kepada kelas yang memiliki jumlah lebih banyak. SVM-RFE unggul dalam *precision* sebesar 99.26% disusul oleh tanpa *feature selection* sebesar 99.16% dan *wrapper method* sebesar 99.15%. Spesificity ketiganya hampir sama yaitu 99.93% untuk SVM-RFE dan 99.92% untuk *wrapper method* dan tanpa *feature selection*.



Gambar 4.5: Grafik akurasi, sensitivity, precision, specificity untuk SVM-RFE, Wrapper method dan tanpa *feature selection*

Tabel 4.1: Perbandingan waktu yang diperlukan untuk learning dan prediksi masing-masing pengklasifikasi dalam detik

	SVM-RFE	Wrapper method	Tanpa feature selection
Train time	4.354	4.764	8.410
Predict time	1.414	1.570	2.599

Tabel 4.1 menunjukkan waktu yang dibutuhkan pada saat *training* pengklasifikasi dan waktu yang dibutuhkan untuk mengklasifikasi 1000 senyawa. Terlihat dengan melakukan *feature selection* waktu yang diperlukan baik untuk *training* maupun prediksi berkurang setengahnya.

Selanjutnya pengklasifikasi SVM-RFE dan pengklasifikasi *wrapper method* digunakan untuk memprediksi senyawa-senyawa mana dalam database senyawa tanaman herbal Indonesia yang dapat bereaksi dengan HIV-1 protease. Pengklasifikasi SVM-RFE mengklasifikasi 368 senyawa yang dapat bereaksi sedangkan

pengklasifikasi *wrapper method* mengklasifikasi 379 senyawa. Hasil prediksi ini dibandingkan dengan penelitian oleh (Yanuar et al., 2014) dan disajikan dalam tabel 4.2. Terlihat pengklasifikasi yang dihasilkan hanya dapat mengklasifikasi 66.7% dari 10 peringkat teratas senyawa pada penelitian (Yanuar et al., 2014).

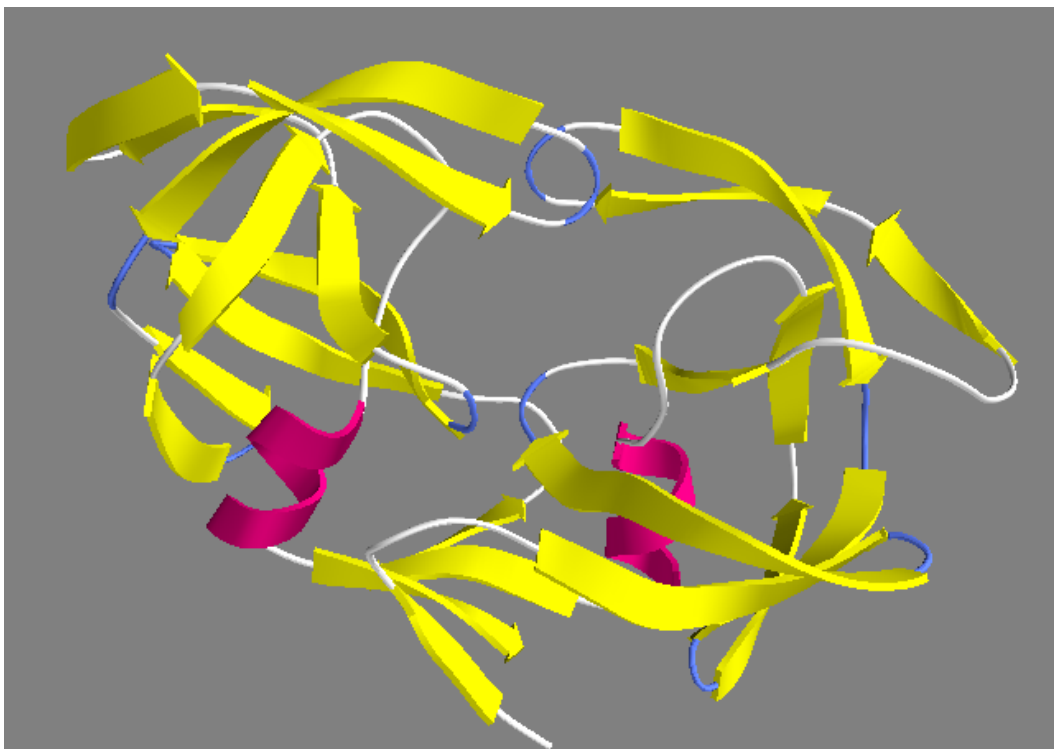
Tabel 4.2: Perbandingan dengan penelitian (Yanuar et al., 2014)

No.	Senyawa Kandidat Obat	SVM-RFE	WM
1.	8-Hydroxyapigenin (2'',4''-disulfatoglucuronide)	Y	Y
2.	Isoscutellarein 4'-methylether 8-(2'',4''-disulfatoglucuronide)	Y	Y
3.	Amaranthin	T	Y
4.	Torvanol A	Y	T
5.	Urosonic acid*	T	T
6.	5-Carboxypyranocyanidin 3-O-(6''-O-malonyl-beta-glucopyranoside)	Y	Y
7.	Oleoside	Y	T
8.	Jacoumaric acid	Y	Y
9.	Platanic acid	T	T
10.	5-Carboxypyranocyanidin 3-O-beta-glucopyranoside	Y	Y

4.4 Molecular Docking

Tahap terakhir adalah melakukan *docking* senyawa-senyawa yang diprediksi sebagai obat oleh pengklasifikasi SVM-RFE pada langkah sebelumnya. Hal ini dilakukan untuk memastikan apakah senyawa-senyawa tersebut dapat bereaksi dengan target obat. *Docking* dilakukan menggunakan PyRx, yaitu *GUI frontend* untuk Autodock.

Target yang digunakan adalah HIV-1 protease yang berasal dari PDB dengan kode 3OCX. Gambar 4.6 menunjukkan struktur target yang digunakan. Sebelum digunakan sebagai target *docking*, protein target dibersihkan dari molekul air dan residu lainnya. Koordinat XYZ yang digunakan pada saat *docking* adalah 5.192, -4.557, 14.799, dengan ukuran *gridbox* 50x50x50 unit, serta evaluasi energi maksimum sebesar 1,000,000.



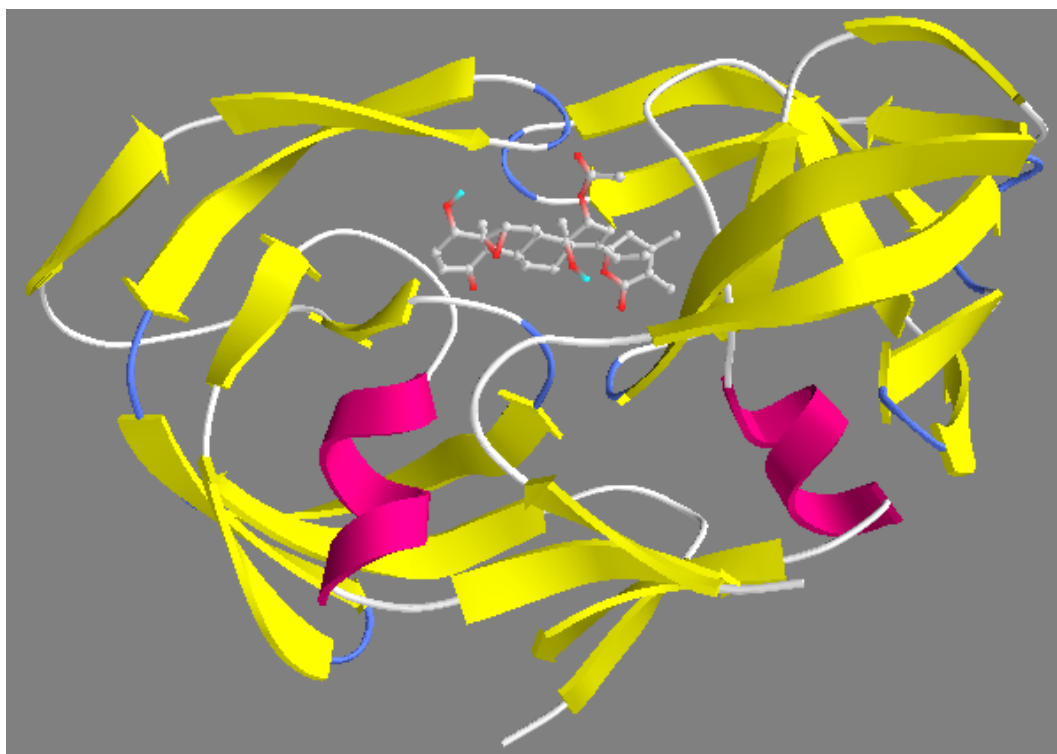
Gambar 4.6: Protein target dalam docking, yaitu HIV-1 protease

Dari 368 senyawa yang didocking, 11 diantaranya tidak dapat bereaksi terhadap target protein. Tabel 4.3 menunjukkan 10 peringkat teratas senyawa hasil docking dengan nilai *bindind energy* terendah. Terlihat beberapa senyawa berasal dari spesies tanaman yang sama. Senyawa Withangulatin A dan Wirtanolide keduanya berasal dari *Physalis Angulata*. Selain itu *Malaleuca Leucadendron* merupakan spesies yang menghasilkan senyawa Leucadenone B, (3beta)-28-Norlup-20(29)-ene-3,17-diol 3-acetate, dan Leucadenone D. Sayangnya senyawa-senyawa dari (Yanuar et al., 2014) tidak masuk 10 besar, mungkin ada perbedaan dalam langkah maupun parameter *docking*. Gambar 4.7 menunjukkan konformasi senyawa withangulatin A dengan protein HIV-1 protease.

Tabel 4.3: Sepuluh senyawa dengan binding energy terendah beserta species asal senyawa tersebut.

No.	Nama Senyawa	Binding Energy	Spesies
1	Withangulatin A	-12.03	<i>Physalis Angulata</i>
2	Withanolide	-11.27	<i>Physalis Angulata</i>

3	Violaxanthin	-11.12	Capsicum Annuum Mangifera Indica Taraxacum Officinale
4	(+)-Thalrugosine	-11.01	Cyclea Barbata
5	Orthosiphonone A	-10.96	Orthosiphon Aristatus
6	Panduratin B	-10.80	Boesenbergia Pandurata
7	Leucadenone B	-10.64	Melaleuca Leucadendron
8	Proanthocyanidin A1	-10.50	Arachis Hypogaea Areca Catechu Cola Acuminata Illicium Anisatum Theobroma Cacao
9	(3beta)-28-Norlup-20(29)-ene-3,17-diol 3-acetate	-10.47	Melaleuca Leucadendron
10	Leucadenone D	-10.45	Melaleuca Leucadendron



Gambar 4.7: Konformasi senyawa Withangulatin A dengan protein HIV-1 protease

BAB 5

PENUTUP

5.1 Kesimpulan

Penelitian ini telah menunjukkan bahwa *feature selection* dapat mempercepat proses *learning* dan prediksi tanpa mengurangi performa pengklasifikasi. Performa awal yang sudah tinggi menyebabkan tidak terlihatnya perbedaan performa yang signifikan antara SVM-RFE, wrapper method, dan tanpa *feature selection*. *Screening* menggunakan LBS pada tahap awal dapat mengurangi senyawa-senyawa yang perlu *discreening* lebih lanjut menggunakan *molecular docking*. Penelitian ini telah melakukan *Virtual screening* terhadap Database Tanaman Herbal Indonesia dan mengidentifikasi 10 senyawa yang bereaksi dengan HIV-1 *protease*.

5.2 Saran

Penelitian ini pada akhirnya tidak luput dari berbagai kekurangan. Berikut adalah beberapa saran-saran untuk penelitian selanjutnya. Pada penelitian ini terlihat akurasi awal telah cukup tinggi sehingga sulit untuk ditingkatkan lagi dengan *feature selection*. Oleh karena itu, penelitian selanjutnya dapat mencoba kembali dengan dataset yang masih rendah akurasi. *Unbalance data* diduga mempengaruhi hasil pengklasifikasi, penelitian selanjutnya dapat mencoba untuk menggali lebih dalam efek *unbalance data* dan cara mengatasinya.

REFERENSI

- Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., ... Overington, J. P. (2014). The chembl bioactivity database: an update. *Nucleic acids research*, 42.
- Chen, J. J. F. & Visco Jr., D. P. (2016). Developing an in silico pipeline for faster drug candidate discovery: virtual high throughput screening with the signature molecular descriptor using support vector machine models. *Chemical Engineering Science*.
- Guyon, I., Weston, J., & Barnhill, S. (2002). Gene selection for cancer classification using support vector machine. *Machine Learning*, 46.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009, November). The weka data mining software: an update. *SIGKDD Explor. Newsl.* 11(1), 10–18. doi:10.1145/1656274.1656278
- Hilman, M. H. (2012). *Analisa teknik data mining dan kinerja infrastruktur komputasi cloud sebagai bagian dari sistem perancangan obat terintegrasi* (Master's thesis, Fakultas Ilmu Komputer Universitas Indonesia).
- Huey, R., Morris, G. M., Olson, A. J., & Goodsell, D. S. (2007). A semiempirical free energy force field with charge-based desolvation. *Journal of Computational Chemistry*.
- Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). Zinc: a free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*, 52.
- Janecek, A. G., Gansterer, W. N., Demel, M. A., & Ecker, G. F. (2008). On the relationship between feature selection and classification accuracy. *JMLR: Workshop and Conference Proceedings*, 4.
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., ... Bryant, S. H. (2016). Pubchem substance and compound databases. *Nucleic Acids Research*, 44.

- Korkmaz, S., Zararsiz, G., & Goksuluk, D. (2014). Drug/nondrug classification using support vector machine with various feature selection strategies. *Computer Methods and Programs in Biomedicine*.
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*.
- Mysinger, M. M., Carchia, M., Irwin, J. J., & Shoichet, B. K. (2012). Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry*, 55(14), 6582–6594. PMID: 22716043. doi:10.1021/jm300687e. eprint: <http://dx.doi.org/10.1021/jm300687e>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Sotriffer, C. A., Flader, W., Winger, R. H., Rode, B. M., Liedl, K. R., & Varga, J. M. (2000). Automated docking of ligands to antibodies: methods and applications. *Methods*. doi:10.1006/meth.1999.0922
- Suhartanto, H., Yanuar, A., Hilman, M., Wibisono, A., & Dermawan, T. (2012). Performace analysis cluster computing envirointments on molecular dynamic simulation of rad gtpase and lox-curcumin molecules with amber. *IJCSI International Journal of Computer Science*.
- Swardiana, I. W. A. (2014). *Modifikasi dan analisis kinerja gpu autodock dalam proses virtual screening senyawa tanaman obat di indonesia sebagai inhibitor enzim-enzim hiv-1* (Master's thesis, Fakultas Ilmu Komputer Universitas Indonesia).
- Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: a review. *Data Classification: Algorithms and Applications*.
- Theodoridis, S. (2015). *Machine learning: a bayesian and optimization perspective*. Jonathan Simpson.
- Wang, Y., Suzek, T., Zhang, J., Wang, J., He, S., Cheng, T., ... Bryant, S. H. (2014). Pubchem bioassay: 2014 update. *Nucleic Acids Research*, 42.

- Wibisono, A. (2012). *Cloud computing model sebagai pendukung proses simulasi dinamika molekuler* (Master's thesis, Fakultas Ilmu Komputer Universitas Indonesia).
- Yanuar, A., Mun'im, A., Lagho, A. B. A., Syahdi, R. R., Rahmat, M., & Suhartanto, H. (2011). Medicinal plants database and three dimensional structure of the chemical compounds from medicinal plants in indonesia. *International Journal of Computer Science Issues*.
- Yanuar, A., Suhartanto, H., Mun'im, A., Anugrah, B. H., & Syahdi, R. R. (2014). Virtual screening of indonesian herbal database as hiv-1 protease inhibitor. *Bioinformation*.
- Yap, C. W. (2010). Padel-descriptor: an open source software to calculate molecular descriptor and fingerprints. *Journal of Computational Chemistry*.

LAMPIRAN

Lampiran 1. Daftar Feature Terpilih

Hanya pada SVM-RFE	Terdapat pada keduanya	Hanya pada Wrapper Method
ATSc1	ALogP	apol
ATSc4	AMR	ATSm2
ATSm5	ATSc2	ATSp3
ATSp2	ATSm1	ATSp4
C1SP2	ATSm3	ATSp5
C3SP3	ATSp1	BCUTc-1h
gmax	BCUTp-1h	BCUTc-1l
HybRatio	BCUTp-1l	BCUTw-1h
Kier2	BCUTw-1l	bpol
Kier3	C1SP1	gmin
LipinskiFailures	C1SP3	Kier1
maxdCH2	C2SP1	maxaaCH
maxHAvin	C2SP2	maxddC
maxHBa	C2SP3	maxdsN
maxHBint5	C3SP2	maxdssC
maxHBint8	C4SP3	maxHBint10
maxHCsats	FMF	maxHBint6
maxHother	fragC	maxHCHnX
maxHsNH2	hmax	maxHdCH2
maxHtCH	hmin	maxHdNH
maxsBr	maxaaaC	maxHsNH3p
maxsCl	maxaaNH	maxHssNH2p
maxsNH2	maxaaO	maxsCH3
maxssCH2	maxdO	maxsI
maxsssCH	maxHaaCH	maxsNH3p
maxsssNHp	maxHaaNH	maxtsC
maxssssC	maxHBd	maxwHBd

Hanya pada SVM-RFE	Terdapat pada keduanya	Hanya pada Wrapper Method
MDEC-12	maxHBint2	McGowan_Volime
MDEC-13	maxHBint3	MDEC-14
MDEN-12	maxHBint7	MDEC-24
MDEN-13	maxHCsatu	MDEO-22
MDEN-22	maxHsOH	minaaaC
MDEN-23	maxHssNH	minaasN
MDEN-33	maxHsssNHp	minHBd
MDEO-11	maxsOH	minHBint9
minaaCH	maxsOm	minHdNH
minaaN	maxssNH2p	minHother
minaaNH	maxssO	minHsNH3p
minaaO	maxsssN	minHsSH
minaasC	maxtN	minHssNH2p
mindCH2	MDEC-11	minsBr
mindsCH	MDEC-34	minsCH3
mindssC	MDEC-44	minsPH2
minHBint10	MDEN-11	minsSH
minHBint5	MDEO-12	minssS
minHCsatu	minaaS	minsssCH
minHtCH	minddssS	minsssN
minsF	mindO	mintCH
minsNH2	mindsN	n3Ring
minsOH	minHaaCH	n6Ring
minssNH2p	minHaaNH	naAromAtom
mintN	minHAvin	naaS
mintsC	minHBa	naasN
MLFER_A	minHBint2	nAcid
MLFER_S	minHBint3	nAtomLC
naaaC	minHBint8	nBonds
naaN	minHCHnX	nC
nAtomLAC	minHCsats	nCl
nBondsD	minHdCH2	ndNH
nBondsT	minHsNH2	ndO
ndsCH	minHsOH	ndsS
ndsN	minHssNH	ndssS

Hanya pada SVM-RFE	Terdapat pada keduanya	Hanya pada Wrapper Method
ndsssP	minHsssNHp	nF
nF11Ring	minsCl	nF7Ring
nFG12Ring	minsOm	nF8Ring
nHAvin	minssCH2	nG12Ring
nHBint6	minssNH	nHBint10
nHCsatu	minsssNHp	nHBint2
nHdsCH	minwHBa	nHBint4
nHsssNHp	MLFER_BO	nHCHnX
nN	MLFER_E	nHdCH2
nP	naaO	nHmisc
nsssCH	naasC	nHsNH3p
nsssNHp	nAtomP	nHsSH
ntN	nBase	nI
nwHBd	ndCH2	nsBr
SaaaC	nddssS	nsCl
SCH-5	nF6Ring	nsNH3p
SdsCH	nHaaNH	nsPH2
SHaaCH	nHBa	nsSH
SHAvin	nHBAcc	nssNH
SHBint2	nHBd	nssssC
SHBint3	nHBDon	nssssSi
SHBint4	nHBint3	nT5Ring
SHBint8	nHBint5	nT6Ring
SHsssNHp	nHBint7	nT9Ring
SP-2	nHBint8	ntCH
SP-3	nHBint9	nTRing
SP-7	nHCsats	SCH-3
SsBr	nHsOH	SCH-7
SsNH2	nHssNH2p	SddssS
SssNH2p	nRotB	SdS
SsssCH	nS	SdssS
StsC	nsOH	SdsssP
VAdjMat	nsOm	SHBd
VC-3	nssNH2p	SHBint10
VP-5	nssO	SHBint6

Hanya pada SVM-RFE	Terdapat pada keduanya	Hanya pada Wrapper Method
VPC-4	nsssN	SHBint7
VPC-5	nT11Ring	SHCsatu
VPC-6	nT8Ring	SHdCH2
WPOL	ntsC	SHdNH
WTPT-2	nwHBa	SHdsCH
	SaaCH	SHmisc
	SaaO	SHsNH3p
	SaasC	SHsOH
	SC-3	SHtCH
	SCH-6	SP-6
	SdCH2	SPC-5
	SdssC	SsCH3
	SHaaNH	SsI
	SHBint5	SsOH
	SHBint9	SsPH2
	SHCsats	SssO
	SHother	SssS
	SHsNH2	SsssNHp
	SP-5	SssssC
	SPC-4	sumI
	SPC-6	SwHBa
	SsCl	SwHBd
	SsOm	VC-5
	SssCH2	VC-6
	SssNH	VCH-3
	SsssN	VP-3
	StN	WPATH
	VC-4	WTPT-1
	VCH-5	WTPT-4
	VCH-6	
	VP-2	
	VP-4	
	VP-7	

Lampiran 2. Daftar Senyawa yang Diklasifikasi sebagai Senyawa Obat

No.	Nama Senyawa	Binding Energy
1	Withangulatin A	-12.03
2	Withanolide	-11.27
3	Violaxanthin	-11.12
4	(+)-Thalrugosine	-11.01
5	Orthosiphonone A	-10.96
6	Panduratin B	-10.80
7	Leucadenone B	-10.64
8	Proanthocyanidin A1	-10.50
9	(3beta)-28-Norlup-20(29)-ene-3,17-diol 3-acetate	-10.47
10	Leucadenone D	-10.45
11	beta-Sitosterol 3-O-beta-D-glucopyranoside	-10.42
12	Isochondodendrine	-10.42
13	Homoaromoline	-10.28
14	Nimolicinol	-10.24
15	6-Deoxoteasterone	-10.23
16	Sapogenin	-10.23
17	Cathasterone	-10.12
18	Tetrandrine	-10.06
19	Camelliaside B	-10.03
20	Mutatoxanthin	-9.98
*21	Jacoumaric acid	-9.93
22	Myricetin 3-(2G-rhamnosylrutinoside)	-9.83
23	Neriifolin	-9.82
24	Heterophyllin	-9.69
25	ent-Epiafzelechin-(4alpha-8)-epicatechin	-9.63
26	Limacine	-9.63
27	Roxburghine B	-9.62
28	Orthosiphol B	-9.60
29	Mangostenone A	-9.60
30	(-)-2-Norlimacine	-9.59
31	Capsanthin 5,6-epoxide	-9.58
32	Methyl angolensate	-9.58

No.	Nama Senyawa	Binding Energy
33	Cyclomorusin	-9.57
34	2''-O-alpha-L-Rhamnosyl-6-C-fucosyl-luteolin	-9.56
35	Azadiradione	-9.56
36	Methyl pheophorbide a	-9.54
37	12-O-2Z,4E-Octadienoyl-4-deoxyphorbol 13-acetate	-9.50
38	Bruceoside A	-9.49
39	Guavacoumaric acid	-9.48
40	Cinchonain Ib	-9.44
41	Typhasterol	-9.43
42	Momordicoside F1	-9.38
43	Teasterone	-9.37
44	Leucadenone A	-9.36
45	Diosgenin	-9.36
46	(-)-Curine	-9.35
47	Leucadenone C	-9.31
48	Dolichosterone	-9.30
49	Guavanoic acid	-9.24
50	Cycloheterophyllin	-9.22
51	(+)-Cycleabarbaine	-9.17
52	Guibourtinidol-(4alpha-8)-epiafzelechin	-9.13
53	Karaviloside I	-9.12
54	Vinblastine	-9.11
55	5,7,2',4'-Tetrahydroxy-6,3'-diprenylisoflavone 5-O-(4''-rhamnosylrhamnoside)	-9.00
56	2''-O-alpha-L-Rhamnosyl-6-C-fucosyl-3'-methoxyluteoin	-8.97
57	Rubranine	-8.97
58	Brassinolide	-8.94
59	Peonidin 3-(4'-arabinosylglucoside)	-8.88
60	Eurycolactone E	-8.87
61	ent-Epiafzeiechin-(4alpha-8)-epiafzelechin	-8.83
62	Cassiaoccidentalin B	-8.82
63	Sanggenol O	-8.82
64	Cycloartocarpin A	-8.81
65	(+)-Cycleanorine	-8.81
66	Kaempferol 3-robinobioside	-8.80

No.	Nama Senyawa	Binding Energy
67	Bryophyllin A	-8.79
68	2''-O-alpha-L-Rhamnosyl-6-C-quinovopyranosyl-luteolin	-8.76
69	Cucurbitacin C	-8.75
70	Euphorbia factor Ti2	-8.74
71	dl-Fangchinoline	-8.67
72	28-Norcastasterone	-8.66
73	Tetrandrine 2'beta-N-oxide	-8.64
74	BR-Xanthone A	-8.58
75	Orthosiphol A	-8.57
76	Neoorthosiphol A	-8.57
77	Caesaldekarine A	-8.57
78	6-Hydroxykaempferol 3,6-diglucoside	-8.55
79	Cycleadrine	-8.55
80	Camarinic acid	-8.52
81	Boesenbergin B	-8.49
82	(-)-Repandine	-8.49
83	Meloside A	-8.41
84	Vitexilactone	-8.40
85	Mesuagin	-8.38
86	Scutellarein 7-neohesperidoside	-8.34
87	Isoforskolin	-8.33
88	Pseudoakuammigine	-8.27
89	Cyanidin 4'-glucoside	-8.25
90	Echitamine	-8.21
91	Corilagin	-8.20
92	Dehydrodihydrorotenone	-8.19
93	Isotetrandrine	-8.17
94	Gibberellin A20	-8.10
95	Carthamin	-8.09
96	Cinchonain Ia	-8.08
97	Kaempferol 3-rhamnosyl- (1-3)(4'''-acetylramnosyl)(1-6)-glucoside	-8.07
98	Kaempferol 3-(3'',4''-diacetylramnoside)	-8.06
99	Gibberellin A38	-8.05
100	Pongachalcone I	-8.02

No.	Nama Senyawa	Binding Energy
101	Quercetin 3-glucosyl-(1-2)-[rhamnosyl-(1-6)-galactoside]	-8.01
102	Rotundifuran	-8.01
103	Sojagol	-8.01
104	Arborside A	-8.00
105	(-)-Blumealactone B	-8.00
106	1-Isomangostin	-8.00
107	Saponarin	-7.99
108	Boesenbergin A	-7.99
109	Coleonol C	-7.98
110	5,7-Dihydroxy-4'-methoxy-8-C-prenyl-3'-(3-hydroxy-3-methylbutyl)flavanone	-7.98
111	Gibberellin A34	-7.96
112	Achillicin	-7.95
113	Proanthocyanidin A2	-7.92
114	alpha-Tocotrienol	-7.91
115	Hydroxycristacarpone	-7.91
116	Idaein	-7.90
117	Cyanidin 3-laminaribioside	-7.90
118	Tinctormine	-7.87
119	Kaempferol 3-rhamnosyl-(1-3)-rhamnosyl-(1-6)-glucoside	-7.87
120	Carlinoside	-7.86
121	Gambiriin C	-7.86
122	Rhamnetin 3-rhamnosyl-(1-4)-rhamnoside	-7.82
123	(-)-Blumealactone A	-7.82
124	Abrussaponin I	-7.79
125	5,2',3'-Trihydroxy-7,8-dimethoxyflavone 3'-glucoside	-7.76
126	Fuzitine	-7.75
127	Isoscoparin 2''-(6-(E)-p-coumaroylglucoside)	-7.73
128	Arborside B	-7.73
129	Cyanidin 3,4'-diglucoside	-7.70
130	Gibberellin A3	-7.70
131	Empetrin	-7.69
132	Epicatechin-(4beta-8)-ent-epicatechin	-7.68
133	Jasmolactone B	-7.68
134	Artonin K	-7.68

No.	Nama Senyawa	Binding Energy
135	Clitorin	-7.67
136	Tetrandrine N-2'-oxide	-7.65
137	Gibberellin A15	-7.65
138	Artonin L	-7.64
139	Peonidin 3-arabinoside	-7.61
140	4-O-Methylsappanol	-7.59
141	Precatorin II	-7.58
142	Paniculonol isovalerate	-7.57
143	Gibberellin A44	-7.56
144	12a-Hydroxyrotenone	-7.55
145	Heliettin	-7.54
146	Kaempferol 3-(6''-acetylglucosyl)-(1-3)-galactoside	-7.53
147	Camelliaside A	-7.52
148	Samarangenin A	-7.47
149	Orientanol A	-7.46
150	gamma-Tocopherol	-7.46
151	Boeravinone C	-7.45
152	Gibberellin A77	-7.45
153	Strigol	-7.44
154	Petunidin 3-glucoside	-7.43
155	Tauremizin	-7.43
156	Gibberellin A29	-7.41
157	(-)-Blumealactone C	-7.37
158	Gibberellin A9	-7.36
159	Cyanidin 3-[6-(6-p-coumarylglucosyl)-2-xylosylgalactoside]	-7.35
160	Scutellarein 6,4'-dimethyl ether 7-(6''-acetylglucoside)	-7.35
161	Epicatechin-(4beta-8)-epiafzelechin	-7.34
162	alpha-Tocopherol	-7.33
163	Gibberellin A55	-7.33
164	Gibberellin A5	-7.33
165	(-)-N-Methylcoclaurine	-7.30
166	Delphinidin 3-O-beta-D-(6-O-(E)-p-coumaryl) galactopyranoside	-7.28
167	delta-Tocopherol	-7.28
168	Gibberellin A98	-7.28

No.	Nama Senyawa	Binding Energy
169	Nicotiflorin	-7.27
170	Loganin	-7.26
171	Rutacridone	-7.26
172	Xanthyletin	-7.26
173	Guanosine	-7.25
174	Gibberellin A1	-7.23
175	Xanthoplanine	-7.22
176	Jasmolactone A	-7.21
177	Hesperidine	-7.20
178	Mirabijalone A	-7.20
179	Gibberellin A51	-7.20
180	Galangin 3,7-dimethyl ether	-7.18
181	Arbortristoside D	-7.16
182	Palmarumycin CP1	-7.15
183	Isovitexin 2''-O-(6'''-(E)-p-coumaroyl)glucoside	-7.14
184	Candidol	-7.14
185	(-)-Cycleapeltine	-7.13
186	Plantamajoside	-7.12
187	Kaempferol 3-glucosyl-(1-3)-rhamnosyl-(1-6)-galactoside	-7.10
188	7,4'-Dihydroxyflavone	-7.10
189	Gibberellin A35	-7.10
190	[Epiafzelechin-(4beta-8)]2-epiafzelechin	-7.09
191	Malvidin	-7.09
192	6-Methoxypulcherrimin	-7.08
193	Isovitexin 2''-O-(6'''-feruloyl)glucoside	-7.05
194	Gibberellin A8	-7.05
195	Asperuloside	-7.04
196	Quercetin 3-rhamnosyl-(1-2)-alpha-L-arabinopyranoside	-7.03
197	Kaempferol 3-(3Rha-glucosylrutinoside)	-7.02
198	Cyanin	-7.02
199	(+)-Marmesin	-7.02
200	beta-Tocopherol	-7.01
201	Allamandicin	-7.01
202	Arborside C	-7.00
203	Cyanidin 3-arabinoside	-6.96

No.	Nama Senyawa	Binding Energy
204	Arbortristoside E	-6.96
205	Cyanidin 3-O-glucoside	-6.95
206	Cyanidin 3-arabinoside	-6.92
207	Gibberellin A4	-6.91
208	Marmin	-6.90
209	Vicenin 2	-6.89
210	Peonidin 3-(6'-malonylglucoside)-5-glucoside	-6.89
211	Allamdin	-6.88
212	Safflor yellow A	-6.87
213	Melampyroside	-6.86
214	Paederoside	-6.85
215	Nirurin	-6.83
216	Gibberellin A7	-6.83
217	Zerumbone epoxide	-6.83
218	Pulcherrimin	-6.79
219	Paniculide A	-6.78
220	Acetylbarlerin	-6.77
221	Mirabijalone D	-6.75
222	Euphorbianin	-6.73
223	Xanthoxyletin	-6.73
224	5-Methoxyseselin	-6.73
225	Quercetin 5,7,3',4'-tetramethyl ether	-6.71
226	Kaempferol 3-neohesperidoside	-6.69
227	[1S-(1alpha,4aalpha,7aalpha)]-7-[(acetyloxy)methyl] -1-(beta-D-glucopyranosyloxy) -1,4a,5,7a-tetrahydrocyclopenta[c]pyran-4-carboxylic acid	-6.69
228	Tricyclohumuladiol	-6.69
229	Oenin	-6.68
230	Baicalein	-6.63
231	Eleutherin	-6.63
232	9-O-Methyl-4-hydroxyboeravinone B	-6.61
233	Vitexin 7-O-glucoside	-6.60
*234	Torvanol A	-6.60
235	3'-Deoxysappanone B	-6.58
236	beta-Lapachone	-6.58

No.	Nama Senyawa	Binding Energy
237	Barlerin	-6.57
238	Verbascoside	-6.56
239	[1aR-(1aalpha,2alpha,4aalpha,7alpha,8aS)] -Octahydro-1a,4a-dimethyl -7-(1-methylethenyl)-3H-naphth[1,8a-b]oxiren-2-ol	-6.56
240	cyanidin-3,5-diglucoside	-6.52
241	Azadirachtin A	-6.48
242	alpha-Rotunol	-6.47
243	Guibourtinidol	-6.44
244	Kandelin B1	-6.40
245	Oxycurcumenol	-6.40
246	Seselin	-6.39
247	Petunin	-6.37
248	Peonin	-6.37
249	(+)-beta-Rotunol	-6.36
250	Paederosidic acid	-6.35
251	Kaempferol 3-(6G-malonylneohesperidoside)	-6.32
252	Rhamnetin 3-mannosyl-(1-2)-alloside	-6.27
253	Allokhusiol	-6.24
254	Cyanidin 3-(6"-malonyllaminaribioside)	-6.21
255	Quercetin 3-2G-rhamnosylrutinoside	-6.20
256	Cyanidin 3-(6"-malonylglucoside)-5-glucoside	-6.20
257	1-Methoxyspirobrassinin	-6.16
258	Lumichrome	-6.16
259	Cyanidin 3-sophoroside-5-glucoside	-6.14
260	Cyanidin 3-(6"-malonylglucoside)	-6.14
*261	5-Carboxypyranocyanidin 3-O-beta-glucopyranoside	-6.12
262	Agmatine	-6.12
263	Delphinidin 3-(2-rhamnosyl-6-malonylglucoside)	-6.11
264	Oxycoccicyanin	-6.11
265	Ipomoeamarone	-6.11
266	5-Carboxypyranocyanidin 3-O-(6"-O-malonyl-beta-glucopyranoside)	-6.10
267	6-Hydroxykaempferol 3,6,7-triglucoside	-6.03
268	Limocitrol 3-neohesperidoside	-6.03

No.	Nama Senyawa	Binding Energy
269	Flavokawin B	-6.03
270	6alpha-Hydroxyadoxoside	-6.01
271	Quercetin 3-sophoroside-7-glucoside	-6.00
272	Paederosidic acid methyl ester	-5.99
273	Psoralen	-5.87
274	Vasicinone	-5.75
275	Delphin	-5.73
276	Isovitexin 2''-O-(6'''-(E)-feruloyl)glucoside 4'-O-glucoside	-5.73
277	eKaempferol 3-O-sophoroside 7-O-beta-D-glucopyranoside (-)-Kaempferol 3-O-sophoroside 7-O-beta-D-glucopyranoside	-5.73
278	Betanidine	-5.69
279	Asimicin	-5.67
280	Omphalocarpin	-5.67
281	Delphinidin 3-(6''-malonylglucoside)	-5.66
282	Sinapoylcholine	-5.63
283	Theviridoside	-5.58
284	Majoroside	-5.57
285	Secoxyloganin	-5.57
286	Casuarinin	-5.56
287	Quercetin 3-glucosyl-(1-2)-galactosyl-(1-2)-glucoside	-5.54
288	Kaempferol 3-glucosyl-(1-2)-galactosyl-(1-2)-glucoside	-5.51
289	Camelliaside C	-5.51
290	Luteolin 7-O-(2-apiofuranosyl-4-glucopyranosyl-6-malonyl) glucopyranoside	-5.50
291	Umbelliferone	-5.49
292	Isovitexin 2''-O-(6'''-(E)-p-coumaroyl)glucoside 4'-O-glucoside	-5.48
293	Murracarpin	-5.48
294	hypaphorine	-5.48
295	Safflomin C	-5.47
296	Asperulosidic acid	-5.44
297	Vasicine	-5.41
298	Abrussaponin II	-5.40
299	Glucobrassicin	-5.33
300	Scopoletine	-5.33
301	Pseudopelletierine	-5.22

No.	Nama Senyawa	Binding Energy
*302	8-Hydroxyapigenin 8-(2'',4''-disulfatoglucuronide)	-5.16
303	Scandoside methyl ester	-5.16
304	Ferulic acid methyl ester	-5.15
305	Peonidin 3-sophoroside-5-glucoside	-5.12
306	Miraxanthin-V	-5.07
307	Sylvestroside I	-5.06
308	Coumarin	-5.05
309	Terpene	-5.05
310	Cantleyoside	-5.04
311	Murrayacarpin A	-5.03
312	2-Phenylethylglucosinolate	-5.02
313	Hydroxysafflor yellow A	-4.95
314	Kaempferol 3-glucosyl -(1-4)(6'''-sinapylglucosyl)(1-2)-galactoside	-4.81
315	rehmannioside D	-4.81
316	Quercetin 3-sophoroside-7-glucuronide	-4.81
317	Multiroside	-4.71
318	Pycnarrhine	-4.69
319	Casuarictin	-4.67
320	Dehydromatricaria ester	-4.67
*321	Isoscutellarein 4'-methyl ether 8-(2'',4''-disulfatoglucuronide)	-4.63
322	Squamocin	-4.57
323	Quercetin 3-(2G-rhamnosylrutinoside)	-4.56
324	6-Hydroxykaempferol 3-rutinoside-6-glucoside	-4.52
325	Shazhiside	-4.48
326	Scopoline	-4.46
327	Caffeine	-4.43
328	Quercetin 3-(6'''-feruloylglucosyl) -(1-2)-galactosyl-(1-2)-glucoside	-4.35
329	Acetovanillone	-4.34
330	Cyanidin 3,5-di-(6-malonylglucoside)	-4.19
331	Indicaxanthin	-4.00
332	Hippuric acid	-3.97
333	Kaempferol 3-sophoroside-7-glucuronide	-3.86
334	Cyanidin 3-(3'',6''-dimalonylglucoside)	-3.85

No.	Nama Senyawa	Binding Energy
335	Trigonelloside C	-3.82
336	Shikimic acid 3-phosphate	-3.48
337	Cyanidin 3-[6-(6-ferulylglucosyl)-2-xylosylgalactoside]	-3.43
338	Cyanidin 3-(6''-caffeylsophoroside)-5-glucoside	-3.34
339	Cyanidin 7-(3-glucosyl-6-malonylglucoside)-4'-glucoside	-3.18
340	Precarthamin	-2.93
341	Choline	-2.80
342	Cyanidin 3-(3-glucosyl-6-malonylglucoside)-4'-glucoside	-2.62
343	Miraxanthin-I	-2.54
344	Betaine	-2.39
345	Punicalin	-2.14
346	Peonidin 3-ferulyldiglucoside-5-glucoside	-2.06
347	Miraxanthin-II	-2.01
348	Cyanidin 3-[6-(6-sinapylglucosyl)-2-xylosylgalactoside]	-1.77
349	[Epicatechin-(4beta-8)]2-epicatechin-(4beta-6)-catechin	-1.46
350	Samarangenin B	-1.02
351	Theasaponin E2	-0.17
352	Theasaponin E1	1.01
353	Theasaponin	1.22
354	Arecatannin A2	1.43
355	Punicalagin	9.93
356	Guajavin B	93.73
357	Eugenigrandin A	142.72
358	Anhydrosafflor yellow B	N/A
359	Capsicoside A	N/A
360	Delphinidin 3-(6-malonylglucoside) -3',5'-di-(6-p-coumaroylglucoside)	N/A
361	Kaempferol 3-(2'''-sinapylglucosyl) (1-4)(6'''-sinapylglucosyl)(1-2)-galactoside	N/A
362	Myristin	N/A
363	Operculinic acid	N/A
364	Safflor yellow B	N/A
365	Woodfordin A	N/A
366	Woodfordin B	N/A
367	Woodfordin C	N/A

No.	Nama Senyawa	Binding Energy
368	Woodfordin D	N/A