

Identification Pharmacodynamic Interactions of Active Compounds of Diabetes Mellitus Type 2 Herbal Plants Using the Random Forest Method*

M. Aiman Askari¹, Farit M. Afendi^{2†}, Anwar Fitrianto³, and Sony Hartono Wijaya⁴

^{1,2,3,4}Department of Statistics, IPB University, Indonesia,
†corresponding author : fmafendi@apps.ipb.ac.id

Copyright © 2022 M. Aiman Askari, Farit M. Afendi, Anwar Fitrianto, and Sony Hartono Wijaya. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Drug-drug interactions is defined as the modification of the effect of a drug as a result of another drug given simultaneously or with an interval or when two or more drugs interact so that the effectiveness or toxicity of one or more drugs changes. Pharmacodynamic interactions are one type of interaction that needs special attention because these interactions work directly on the body's physiological systems and compete on the same receptors so that they can be antagonistic, additive, or synergistic. The use of medicinal plants is becoming an alternative because in addition to their relatively safer side effects, medicinal plants consisting of active compounds are appropriate in treating degenerative metabolic diseases triggered by mutations in many genes. As in the case of polypharmacies, interactions of active compounds in medicinal plants can also lead to pharmacodynamic interactions. Therefore, it is also necessary to identify the active compounds so that it can then be known whether the interaction of the compounds will be beneficial or detrimental. In this study, pharmacodynamic identification was applied to Diabetes Mellitus Type 2 medicinal plant compounds by using the independent variables Target Protein Connectedness (TPC), Side Effect Similarity (SES), and Chemical Similarities (CS) using Random Forest classification method. From a search of various databases, 21 active compounds were obtained and then only 100 compound interactions could be calculated as independent variables. With an accuracy value and AUC of 0,96, there were 93 pairs of compounds that interacted pharmacodynamically and the remaining 7 did not interact.

Keywords: chemical similarities, pharmacodynamic interaction, random forest, side effect similarity, target protein connectedness.

* Received: Aug 2021; Reviewed: May 2022; Published: Aug 2022

1. Pendahuluan

Interaksi obat-obat adalah suatu interaksi yang terjadi pada penggunaan konsumsi lebih dari satu obat dalam proses penyembuhan suatu penyakit atau polifarmasi. Interaksi obat-obat didefinisikan sebagai modifikasi efek suatu obat karena obat lain yang diberikan bersamaan atau pada interval waktu atau ketika dua atau lebih obat berinteraksi sehingga efektivitas atau toksisitas suatu obat atau lebih berubah (Merle *et al.* 2005). Interaksi yang terjadi pada obat dikelompokkan menjadi tiga jenis yaitu farmasetik, farmakokinetik (PK), dan farmakodinamik (PD). Interaksi farmasetik terjadi karena ketidakcocokan fisik atau kimiawi antar obat. Interaksi farmakokinetik terjadi ketika suatu obat mengubah penyerapan, distribusi, metabolisme, atau ekskresi obat lain dan mengubah konsentrasi obat di lokasi sasaran. Interaksi farmakodinamik terjadi jika suatu obat memiliki sifat antagonis, aditif, sinergis, atau efek farmakologis tidak langsung yang lain (Huang *et al.* 2013). Interaksi farmakodinamik bekerja langsung pada sistem fisiologis tubuh dan berkompetisi pada reseptor yang sama sehingga efek suatu obat dapat berubah karena dipengaruhi oleh obat lain. Identifikasi Interaksi farmakodinamik perlu dilakukan agar selanjutnya dapat diketahui apakah kedua obat yang berinteraksi tersebut akan berdampak baik atau justru menghambat proses penyembuhan.

Konsumsi obat untuk proses penyembuhan suatu penyakit dalam jangka panjang akan menimbulkan efek samping yang terakumulasi dan dapat merugikan kesehatan. Untuk mengantisipasi hal tersebut, pemanfaatan tanaman obat menjadi salah satu alternatif. Meskipun digunakan dalam jangka waktu yang lama, efek samping pada tanaman obat lebih sedikit sehingga relatif lebih aman. Keuntungan lain dari obat berbasis tanaman atau jamu yaitu kemampuannya untuk memiliki lebih dari satu efek farmakologi karena terdiri dari banyak senyawa aktif. Sehingga tanaman obat lebih tepat dalam menangani penyakit metabolik degeneratif yang dipicu oleh mutasi pada banyak gen (Katno dan Pramono 2008).

Seperti halnya pada interaksi obat-obat, tahapan penyembuhan menggunakan tanaman obat yang melibatkan banyak komponen senyawa aktif yang bekerja pada satu atau beberapa target juga dapat menyebabkan terjadinya interaksi farmakodinamik. Evaluasi interaksi farmakodinamik pada tanaman obat dapat dilakukan secara eksperimental untuk meramu senyawa-senyawa aktif yang dapat meningkatkan efektivitas penyembuhan. Namun hal ini membutuhkan biaya dan tenaga yang besar serta waktu yang lama karena dilakukan secara kasus per kasus. Oleh karena itu, pendekatan komputasi muncul sebagai suatu alternatif untuk dapat mengidentifikasi interaksi senyawa aktif secara lebih cepat dan efisien.

Penelitian terkait interaksi obat-obat telah dilakukan oleh beberapa peneliti sebelumnya. Huang *et al.* (2013) mengukur kekuatan *Target Protein Connectedness* (TPC) antar obat dan menggunakan *Side Effect Similarity* (SES) dengan menerapkan pendekatan bayes untuk memprediksi interaksi obat. (Mulia 2017) memprediksi senyawa protein menggunakan fungsi *Chemical Similarities* (CS) dengan *fingerprint Klekota-Roth* dan menghasilkan fungsi kemiripan Tanimoto memiliki nilai akurasi dan nilai *Area Under Curve* (AUC) yang terbaik. AUC merupakan kuantifikasi performa dari suatu pengujian berupa luas daerah di dalam grafik *Receiver Operating Characteristic* (ROC) yang merepresentasikan titik seimbang (*tradeoff*) relatif antara keuntungan (*true positive*) dan biaya (*false negative*) (Fawcett 2006). Kemudian

Hasnita *et al.* (2020) dan Winata *et al.* (2019) memprediksi interaksi farmakodinamik pada obat menggunakan peubah penjelas TPC, SES, dan CS. Hasnita *et al.* (2020) membandingkan model *Random Forest*, *Support Vector Machine* (SVM), dan Regresi Logistik Biner dengan menggunakan gugus data obat dari *DrugBank* berupa interaksi farmakodinamik (*Golden Standard Positive/GSP*) dan tidak berinteraksi (*Golden Standard Negative/GSN*). Penelitian ini menghasilkan model SVM dengan nilai kebaikan terbaik. Winata *et al.* (2019) mengelaborasi dataset obat untuk menemukan GSN yang benar-benar tidak berinteraksi menggunakan algoritma *Density-Property Cluster* (DP-Clust). DP-Clust dikembangkan oleh Amin *et al.* (2007) dengan mengelompokkan grafik berdasarkan kepadatan-pinggiran untuk mendeteksi kompleks protein dalam jaringan interaksi.

Selanjutnya pada penelitian ini, identifikasi farmakodinamik akan diterapkan pada tanaman jamu guna pengobatan salah satu penyakit metabolik degeneratif yaitu Diabetes Melitus Tipe 2 menggunakan informasi keterhubungan protein target (TPC), kesamaan efek samping (SES), dan kesamaan struktur kimia (CS) pada senyawa aktif. Tanaman jamu Diabetes Melitus Tipe 2 yang digunakan adalah pare, sembung, bratawali, dan jahe berdasarkan penelitian Nurishmaya (2014). Nurishmaya (2014) menggunakan pendekatan bioinformatika dengan ikan Zebra (*Danio rerio*) sebagai hewan model serta ramuan jamu yang sedang dikembangkan di Pusat Studi Biofarmaka Tropika Bogor dan menemukan bahwa 4 tanaman yaitu pare (*Momordica charantia*), sembung (*Blumea balsamifera*), bratawali (*Tinospora crispa*), dan jahe (*Zingiber officinale*) mampu menurunkan kadar gula darah dibandingkan 20 kelompok tanaman lainnya pada ikan Zebra. Gugus data model yang digunakan merupakan data obat berinteraksi farmakodinamik (GSP) yang telah dikumpulkan Hasnita *et al.* (2020) dan Winata *et al.* (2019) dengan menggunakan *semi-automatic text mining* dari *DrugBank* dan hasil pengelompokan GSN terbaik dari Winata *et al.* (2019). Kemudian model klasifikasi terbaik *Random Forest* akan digunakan untuk mengidentifikasi kelas interaksi farmakodinamik senyawa aktif pada tanaman jamu Diabetes Melitus Tipe 2.

2. Metodologi

2.1 Data

Dalam pembuatan model klasifikasi, gugus data yang digunakan merupakan kombinasi data obat yang berinteraksi farmakodinamik (PD/GSP) hasil pengolahan data yang dilakukan oleh Hasnita *et al.* (2020) dan Winata *et al.* (2019) menggunakan metode *semi-automatic text mining* dan data obat yang tidak berinteraksi (GSN) yang diolah oleh Winata *et al.* (2019) menggunakan metode DP-Clust. GSP dicari melalui metode *semi-automatic text mining* dengan cara mengeliminasi interaksi farmasetik (mengandung salah satu kata kunci "*physicochemical*", "*non-absorbable*", dan "*solution*") dan interaksi farmakokinetik (mengandung salah satu kata kunci "*CYP*", "*distribution*", "*absorption*", "*metabolism*", "*concentration*", "*excretion*", "*level*", "*metabolite*", "*metabolized*", "*clearance*", "*enzyme inducer*"). Data pasangan obat yang tidak mengandung salah satu kata kunci di atas (baik kata kunci untuk interaksi farmasetik maupun farmakokinetik) merupakan interaksi farmakodinamik. Data GSN yang merupakan pasangan obat yang tidak memiliki interaksi dibangun dari metode DP-Clust dengan *tuning* parameter optimum yaitu *cluster* = 37, *cluster property* (cp)=

0,5, *density* (d) = 0,9 sehingga menghasilkan nilai kebaikan model AUC terbaik 0,610 (Winata et al. 2019). Total data yang digunakan sebagai data model yaitu 11.736 (10.546 GSP dan 1.190 GSN).

Kemudian data model akan diterapkan pada senyawa aktif yang terdapat dalam tanaman untuk pengobatan Diabetes Melitus Tipe 2 yang terdiri dari pare (*Momordica charantia*), sembung (*Blumea balsamifera*), bratawali (*Tinospora crispa*), dan jahe (*Zingiber officinale*). Data senyawa aktif diperoleh dari pangkalan data *Knapsack Family, Indonesia Jamu–Herbs (IAH) Analytics*, dan *Dictionary of Natural Products*. Selanjutnya akan dihitung peubah penjelas berupa *Target Protein Connectedness* (TPC), *Side Effect Similarities* (SES), dan *Chemical Similarities* (CS) berdasarkan informasi dari senyawa aktif tanaman untuk diidentifikasi interaksi farmakodinamikanya.

2.2 Prosedur Analisis Data

Prosedur analisis data yang akan dilakukan adalah sebagai berikut:

1. Mengumpulkan data obat yang berinteraksi (GSP) sebagai data dalam pembuatan model klasifikasi sebanyak 10.546.
2. Mengumpulkan data obat tidak berinteraksi (GSN) menggunakan algoritma *Density-Property Cluster* dengan parameter optimal $cluster = 37$, $cp = 0,5$, $d = 0,9$ sebagai data dalam pembuatan model klasifikasi sebanyak 1.190 data GSN.
3. Membuat peubah baru (interaksi peubah) dengan cara mengalikan setiap peubah.
4. Melakukan metode *oversampling* untuk menangani kelas data pasangan obat yang tidak seimbang dengan menambah data GSN sebagai kelas data minoritas agar seimbang dengan jumlah kelas data GSP dengan teknik *resampling*.
5. Membuat model untuk mengklasifikasikan pasangan obat yang berinteraksi farmakodinamik (GSP) dan tidak berinteraksi (GSN) menggunakan metode *Random Forest* dengan menggunakan *5 fold validation* dan kombinasi parameter hingga diperoleh nilai akurasi dan AUC terbaik (*tuning parameter*).
6. Mencari senyawa aktif dengan target khusus terhadap manusia dan protein target yang terdapat dalam tanaman obat Diabetes Melitus Tipe 2.
7. Menghitung nilai TPC sebagai peubah penjelas x_1 . Nilai TPC dihitung berdasarkan informasi bobot protein (*weight score*) yang diperoleh dari *Human Protein Reference Database* (HPRD) (<https://www.hprd.org/>). Data dari HPRD dapat juga digunakan untuk mengetahui pasangan senyawa mana saja yang saling berinteraksi dengan cara melakukan proses *mapping* yang kemudian dihitung dengan persamaan berikut:

$$TPC = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

dimana \bar{x} merupakan nilai rata-rata, s merupakan nilai simpangan baku, n merupakan banyaknya *Pearson Correlation Coefficient* (PCC) ekspresi jaringan-silang dari data yang menghubungkan dua sistem yang berpusat pada senyawa, dan μ_0 merupakan nilai rata-rata PCC pada seluruh data dalam jaringan.

8. Menghitung nilai SES sebagai peubah penjelas x_2 dengan menghitung jarak kesamaan efek samping senyawa. Data efek samping diperoleh dari data *Adverse Drug Reaction* (ADR) pada *website VigiAcces* (<http://www.vigiaccess.org/>) yang kemudian dilakukan penyaringan data untuk menemukan data yang benar-benar

merupakan efek samping di pangkalan data *Side Effect Resource* (SIDER). Perhitungan nilai SES menggunakan persamaan berikut:

$$SES = \sum_{i \in X \cap Y} (-\log f_i) \cdot g_i$$

dimana X dan Y merupakan dua senyawa yang berbeda, f_i merupakan frekuensi efek samping, dan g_i merupakan nilai bobot pada efek samping yang berkorelasi.

- Menghitung jarak kemiripan CS sebagai peubah penjelas x_3 . Nilai CS dihitung menggunakan koefisien *Tanimoto* yang dibangkitkan melalui *fingerprint Klekota Roth* melalui aplikasi web *ChemDes* (<https://www.scbdd.com/chemdes/>). Aplikasi web *ChemDes* digunakan untuk menerima rangkaian *SMILES* yang berkesesuaian dengan struktur kimia dari senyawa pada pembangkitan data *fingerprint*. Perhitungan koefisien *Tanimoto* menggunakan persamaan berikut:

$$CS = \frac{a}{a + b + c}$$

dimana a , b , dan c merupakan frekuensi kemunculan dari kejadian ($x = 1$ dan $y = 1$), ($x = 1$ dan $y = 0$), dan ($x = 0$ dan $y = 1$) untuk tiap bit pada rangkaian dua pasang senyawa.

- Membuat peubah penjelas baru dengan menginteraksikan setiap peubah TPC, SES, dan CS.
- Menentukan kelas pasangan senyawa aktif menggunakan nilai-nilai TPC, SES, dan CS menggunakan model pada langkah (5).

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Pengumpulan data diperoleh dari penelusuran berbagai pangkalan data yang relevan. Pengumpulan informasi ketiga variabel yang digunakan diperoleh dari pangkalan data yang berbeda-beda. Tabel 1 menunjukkan informasi berbagai jenis data, pangkalan data yang menyediakan informasi data tersebut, beberapa informasi yang diperoleh dari pangkalan data (terutama informasi yang diperlukan dalam metode), dan hasil pencarian pada pangkalan data.

Tabel 1: Informasi penelusuran data.

Jenis Data	Pangkalan Data	Informasi yang diperoleh dari Pangkalan Data	Hasil Penelusuran Data
Senyawa	<i>KNAPSAcK family, Indonesia Jamu-Herbs (IJAH) Analytics</i> dan <i>The Dictionary of Natural Products</i> (DNP)	Daftar senyawa alami yang terdapat pada tanaman tertentu (Afendi et al., 2012; Chapman 2008)	250 senyawa pada 4 tanaman yaitu 102 pada pare, 18 pada sembung, 6 pada bratawali, dan 124 pada jahe
Protein target	<i>PubChem BioAssay</i>	Informasi rinci mengenai protein target beserta organismenya (Wang et al. 2009)	Dari 250 senyawa, hanya 97 senyawa yang memiliki protein target terhadap manusia

<i>Adverse Drug Reaction</i> (ADR)	<i>VigiAccess</i>	Informasi <i>Adverse Drug Reaction</i> (ADR) dari otoritas obat nasional di lebih dari 110 negara <i>World Health Organization</i> (WHO)	dengan total 9.201 protein target Diperoleh 21 senyawa dan 2.186 ADR dari 97 senyawa tanaman obat pada pangkalan data <i>VigiAccess</i>
Efek samping senyawa	<i>Side Effect Resource</i> (SIDER)	Informasi data efek samping obat (Campillos et al. 2008)	Dari 2.186 ADR pada pangkalan data <i>VigiAccess</i> , diperoleh 594 efek samping
Struktur kimia senyawa	<i>PubChem</i>	Informasi rangkaian <i>Canonical SMILES</i> senyawa (Wang et al. 2009)	21 struktur kimia senyawa dalam bentuk <i>Canonical SMILES</i>
<i>Fingerprint</i> senyawa	<i>Chemdes</i>	Integrasi struktur <i>Canonical SMILES</i> senyawa ke rangkaian bit <i>fingerprint</i> (Dong et al. 2015)	Rangkaian biner 4.860 fitur <i>Klekota Roth</i> dari 21 struktur senyawa

3.2 Senyawa Aktif Tanaman Jamu

Setelah penelusuran senyawa aktif dengan seluruh informasi yang diperlukan dari berbagai sumber pangkalan data, terdapat 21 senyawa aktif dari tanaman jamu Diabetes Melitus Tipe 2 yang memiliki semua informasi yang dibutuhkan dalam pencarian peubah penjas TPC, SES, dan CS. Senyawa-senyawa aktif itu tersebut adalah *Quercetin*, *Benzoic acid*, *Lycopene*, *beta-Carotene*, *Zeaxanthin*, *Chlorogenic acid*, *beta-Sitosterol*, *Squalene*, *Lutein*, *Rutin*, *Melatonin*, *Eucalyptol*, *Camphor*, *Geraniol*, *Ascorbic acid*, *Genistein*, *Curcumin*, *Linalool*, *Methylglyoxal*, *Propyl alcohol*, dan *Glafenine*.

3.3 Perhitungan Nilai TPC

Nilai TPC dihitung dengan menggunakan informasi protein target terhadap manusia dan jejaring interaksi protein-protein dari 21 senyawa hasil irisan pada penelusuran data untuk ketiga peubah penjas. Protein target diperoleh dari informasi aktifitas biologis pada data *BioAssay* yang terdapat di pangkalan data *Pubchem*. Protein target yang terdapat pada senyawa tidak semuanya merupakan protein target terhadap manusia sehingga perlu diidentifikasi lagi yang benar-benar spesifik merupakan protein target manusia. Tiap protein target memiliki ID unik yang disebut dengan *Target Entrez ID*. *Genistein* merupakan senyawa dengan protein target terbanyak terhadap manusia yaitu 491 protein target, kemudian *Curcumin* 405 protein target, dan *Melatonin* 363 protein target.

Dalam perhitungan TPC, tidak semua pasangan senyawa memiliki nilai TPC. Hal ini disebabkan tidak semua senyawa memiliki minimal satu protein target yang

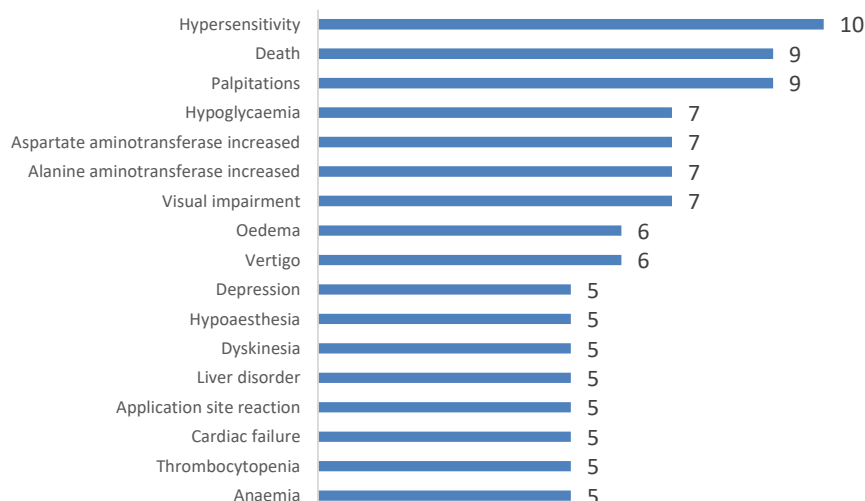
sama sehingga hanya pasangan senyawa yang memiliki minimal satu protein target yang sama yang dapat dihitung nilai TPC-nya.

Setelah informasi protein target manusia dan jaringan interaksi protein-protein diperoleh, langkah selanjutnya yaitu menghitung nilai TPC menggunakan Bahasa pemrograman *Pearl* yang dikembangkan oleh Huang *et al.* (2013). Dari 21 senyawa, diperoleh 100 pasangan senyawa yang berbagi protein target yang sama. 100 pasangan senyawa inilah yang kemudian menjadi pasangan senyawa final yang akan diprediksi interaksinya karena tidak ada lagi penyeleksian pasangan senyawa pada peubah penjelas lainnya.

3.4 Perhitungan Nilai SES

Efek samping senyawa dapat diperoleh di pangkalan data *DrugBank* dengan informasi kode *Anatomical Therapeutic Chemical (ATC)* yang kemudian dapat dicari efek sampingnya di pangkalan data *Side Effect Resource (SIDER)*. Namun, dari 97 senyawa yang terdapat di tanaman pare, sembung, bratawali, dan jahe yang memiliki protein target terhadap manusia, tidak ada satupun yang memiliki informasi kode ATC sehingga tidak dapat diketahui efek sampingnya di pangkalan data *SIDER*. *VigiAcces* merupakan suatu website yang dapat menelusuri dan melihat data tentang dugaan efek samping dari berbagai produk obat yang juga dikenal sebagai *Adverse Drug Reaction (ADR)*. Semua data yang terdapat di *VigiAcces* bersumber dari *VigiBase*, pangkalan data ADR global *World Health Organization (WHO)*, yang dikelola *Uppsala Monitoring Centre (UMC)* yang berbasis di Uppsala, Swedia. Data tersebut berisi laporan dugaan ADR yang dikumpulkan oleh otoritas obat nasional di lebih dari 110 negara. Dari 97 senyawa, hanya 21 senyawa yang memiliki data efek samping di *VigiAcces*.

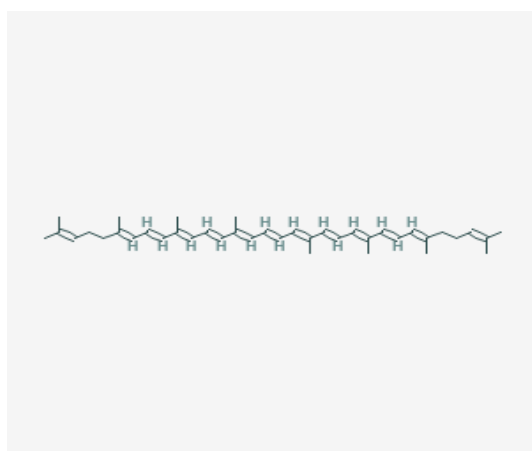
Dari 21 senyawa ini terdapat 2.186 ADR yang kemudian dilakukan proses penyaringan untuk memastikan data yang benar-benar merupakan efek samping yang terdapat di *SIDER*. Proses penyaringan ini menghasilkan 594 data ADR yang benar-benar merupakan efek samping dan terdapat di *SIDER*. *Hypersensitivity*, *Palpitations*, dan *Death* merupakan efek samping terbanyak dari tanaman obat ini. Kemudian untuk menghitung nilai SES, pertama-tama yaitu mencari nilai bobot dari efek samping menggunakan *The Gerstein–Sonnhammer–Chothia algorithm* atau dikenal dengan pembobotan GSC. Pembobotan GSC digunakan untuk menghitung bobot berdasarkan pengelompokan berhirarki matriks jarak *Tanimoto/Jacquard* pada efek samping senyawa. Nilai bobot GSC (g_i) kemudian dimasukkan ke dalam persamaan langkah analisis (8) untuk kemudian mendapatkan nilai SES dari tiap pasangan senyawa.



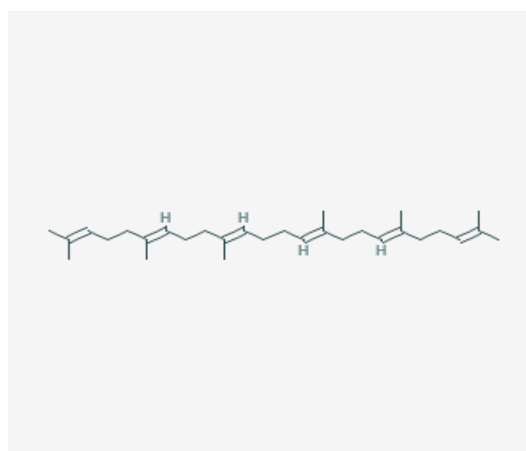
Gambar 1: Frekuensi kemunculan efek samping dari senyawa aktif tanaman jamu Diabetes Melitus Tipe 2 (<http://www.vigiaccess.org>).

3.5 Perhitungan Nilai CS

Nilai *Chemical Similarity* (CS) merupakan kesamaan struktur molekul senyawa yang diperoleh dari integrasi ID *Chemical Abstracts Service* (ID CAS) yang merupakan ID senyawa dari pangkalan data Knapsack, DNP, dan IJAH ke *Chemical ID* (CID) yang merupakan ID senyawa di pangkalan data *Pubchem*. Kemudian struktur senyawa tanaman diperoleh dari *Canonical SMILES* yang kemudian di konversi melalui aplikasi web *Chemdes* untuk mendapatkan *Fingerprint Klekota-Roth*. Selanjutnya *Fingerprint Klekota-Roth* dikonversi menggunakan program *Python* untuk mendapatkan rangkaian *binary digit* (bit) sepanjang 4.860 fitur. Setiap pasangan senyawa kemudian dihitung kemiripan strukturnya berdasarkan rangkaian bit menggunakan rumus kesamaan Tanimoto. Rentang nilai pada rumus kesamaan *Tanimoto* berada antara 0 dan 1 dimana semakin mirip molekul senyawa maka nilai *Tanimoto*-nya akan semakin mendekati 1 dan begitu juga sebaliknya. Senyawa *Lycopene* dan *Squalene* memiliki nilai *Tanimoto* tertinggi yaitu 0,923.



A



B



Gambar 2: Perbandingan struktur molekul senyawa *Lycopene* dan *Squalene*.
(A) Struktur 2D *Lycopene* (B) Struktur 2D *Squalene* (C) Struktur makromolekul *Lycopene* (D) Struktur makromolekul *Squalene*.

Tabel 2 memberikan informasi deskripsi data peubah penjelas yang digunakan dalam analisis. Nilai TPC dan SES memiliki *range* data yang jauh. Sedangkan *range* nilai CS berada diantara 0 dan 1. Semakin tinggi nilai TPC mengindikasikan semakin kuatnya hubungan antara dua senyawa berdasarkan target proteinnya. Semakin banyak kesamaan efek samping antara dua senyawa maka semakin besar juga nilai SES-nya. Semakin mirip struktur molekul senyawa maka nilai CS-nya akan mendekati 1, begitu juga sebaliknya. Tabel 3 menampilkan nilai korelasi antar peubah. Jika dilihat hubungan korelasi antara masing-masing peubah, ketiga peubah penjelas memiliki hubungan korelasi yang sangat lemah.

Tabel 2: Deskripsi nilai TPC, SES dan CS data senyawa aktif tanaman.

	TPC	SES	CS
jumlah	100	100	100
rata-rata	8,299	136,871	0,209
simpangan baku	4,802	72,993	0,154
minimum	0,002	10,023	0,024
kuartil 1	4,595	95,678	0,113
kuartil 2	7,181	133,537	0,181
kuartil 3	12,233	172,042	0,252
maksimum	17,879	274,842	0,923

Tabel 3: Korelasi antar peubah data senyawa aktif tanaman.

	TPC	SES	CS
TPC	1	0,188	-0,157
SES	0,188	1	-0,237
CS	-0,157	-0,237	1

3.6 Pembuatan Model Klasifikasi

Pembuatan model klasifikasi menggunakan data obat yang telah dikumpulkan oleh Hasnita *et al.* (2020) dan Winata *et al.* (2019) dengan total 11.736 pasangan senyawa (10.546 pasangan senyawa berinteraksi farmakodinamik (GSP) dan 1.190 non interaksi (GSN)).

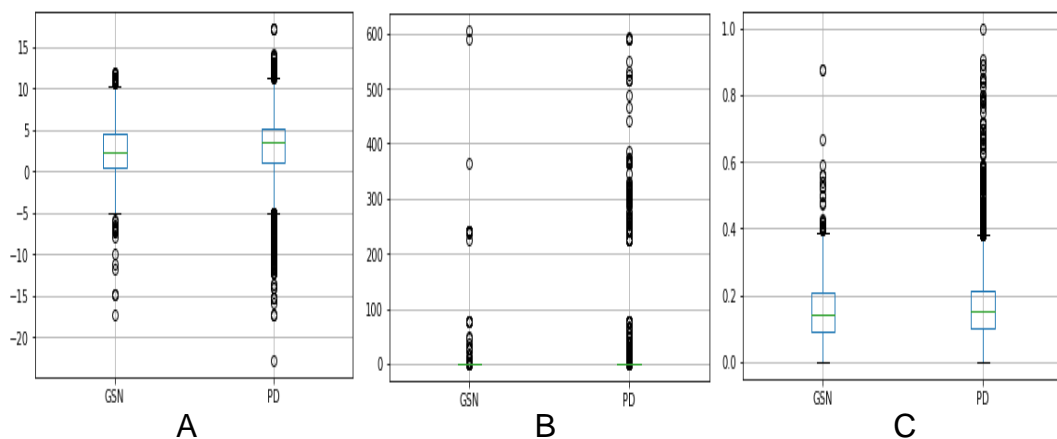
Tabel 4: Deskripsi nilai TPC, SES dan CS data obat.

	TPC	SES	CS
jumlah	11.736	11.736	11.736
rata-rata	2,922	7,559	0,168
simpangan baku	3,575	45,121	0,097
minimum	-22,781	0	0
kuartil 1	1,066	0	0,1
kuartil 2	3,411	0	0,154
kuartil 3	5,151	0	0,214
maksimum	17,320	606,567	1

Tabel 5: Korelasi antar peubah data obat.

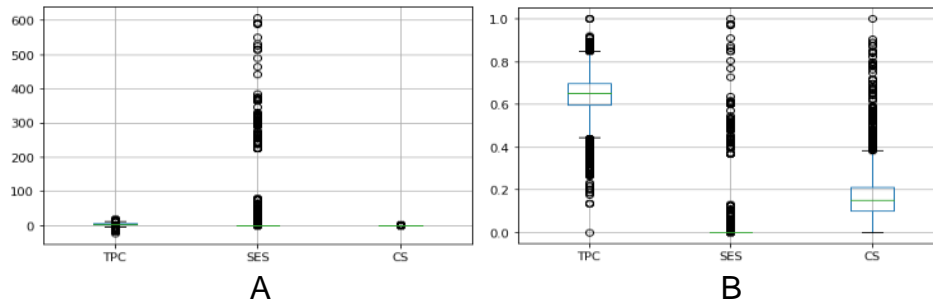
	TPC	SES	CS
TPC	1	0,032	0,014
SES	0,032	1	-0,002
CS	0,014	-0,002	1

Berdasarkan Tabel 4 dan Tabel 5, *range* peubah TPC dan SES memiliki sebaran yang jauh dibandingkan peubah CS, serta korelasi antara ketiga peubah juga sangat lemah.



Gambar 3: Plot perbandingan sebaran nilai data model pasangan obat berinteraksi farmakodinamik (PD/GSP) dan tidak berinteraksi (GSN) pada peubah (A) TPC; (B) SES; dan (C) CS.

Berdasarkan Gambar 3, nilai TPC pada interaksi PD memiliki nilai *median* yang sedikit lebih besar daripada nilai *median* pada kelas GSN. Pada kelas PD juga memiliki sebaran data yang lebih lebar daripada sebaran data pada kelas GSN. Nilai SES pada interaksi PD dan kelas GSN memiliki sebaran data yang relatif sama. Sementara itu nilai CS pada interaksi PD memiliki nilai *median* yang sedikit lebih besar daripada nilai *median* pada kelas GSN. Pada kelas PD juga memiliki sebaran data yang lebih lebar daripada sebaran data pada kelas GSN.



Gambar 4: Plot perbandingan peubah penjelas data model pasangan obat (A) sebelum normalisasi data; (B) sesudah normalisasi data.

Sebaran data nilai TPC dan SES yang cukup jauh dikhawatirkan akan menyebabkan kesalahan dalam memprediksi kelas data dalam model klasifikasi. Untuk menangani hal ini, dilakukan proses normalisasi data menggunakan *MinMax Scaling* untuk membuat rentang data peubah TPC dan SES berada diantara 0 dan 1 (sama dengan rentang data peubah CS).

Dalam pembuatan model klasifikasi *Random Forest*, berbagai tahapan persiapan dan perlakuan data dilakukan untuk menghasilkan model yang terbaik. Tahapan yang dilakukan yaitu membuat interaksi peubah penjelas untuk meningkatkan nilai kebaikan model dan penanganan data *imbalance* berupa *oversampling* agar kelas data menjadi seimbang sehingga meminimalisir kesalahan dalam prediksi kelas data. Kemudian *tuning* parameter dilakukan untuk mengoptimalkan kebaikan model yang diukur dengan nilai akurasi dan AUC. *Tuning* dilakukan pada parameter penggunaan pemilah, jumlah pohon, fitur pemilah, kedalaman pohon maksimum, jumlah minimum sampel yang diperlukan untuk membagi simpul internal, jumlah minimum sampel yang diperlukan untuk berada pada simpul daun, dan penggunaan *bootstrap*. Sementara itu parameter lainnya menggunakan *mode default Random Forest Classifier* pada modul *sklearn* perangkat lunak *Python*. Pada penelitian ini digunakan *5-fold validation* untuk menjaga kestabilan kebaikan model dan nilai kebaikan model *Random Forest* terbaik diperoleh dengan akurasi dan AUC sebesar 0,963.

3.7 Variable Importance

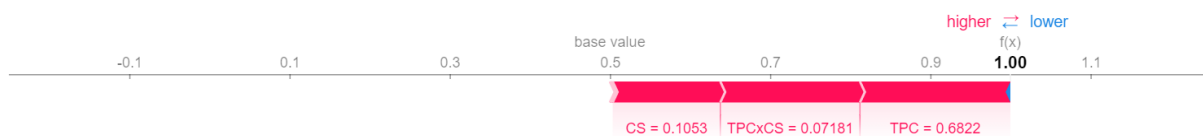
Permutation Importance adalah algoritma untuk mendapatkan informasi *Feature Importance* dengan melakukan permutasi (menyusun ulang kumpulan data) di peubah yang digunakan dalam melatih model prediksi. Berdasarkan Gambar 5, TPC merupakan peubah yang paling berpengaruh. TPC dapat mempengaruhi performa nilai AUC sebesar 0,3478 dengan dilakukan permutasi data dengan variasi sebesar 0.0119. Selain itu peubah berikutnya yang memberikan pengaruh penting terhadap klasifikasi kelas yaitu peubah TPCxCS (interaksi TPC dan CS) dan peubah CS.

Weight	Feature
0,3478 ± 0,0119	TPC
0,1576 ± 0,0031	TPCxCS
0,0917 ± 0,0039	CS
0,0060 ± 0,0005	TPCxSESxCS
0,0059 ± 0,0004	SESxCS
0,0056 ± 0,0005	TPCxSES
0,0046 ± 0,0003	SES

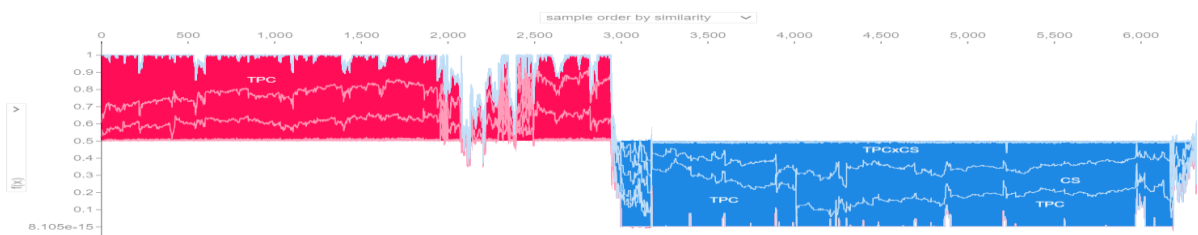
Gambar 5: Tingkat kepentingan peubah penjelas berdasarkan *Permutation Importance*.

Shapley Additive exPlanations (SHAP) adalah sebuah algoritma untuk melihat bagaimana setiap peubah secara spesifik mempengaruhi setiap prediksi yang dibuat. Algoritma ini akan menghitung nilai SHAP dari setiap peubah terhadap prediksi yang dilakukan. Nilai yang dihasilkan merupakan seluruh perhitungan SHAP untuk setiap kelas. Peubah penjelas itu sendiri akan menghasilkan nilai dasar dimana secara umum merupakan rata-rata *output* model berdasarkan data *training* yang dimasukkan. Nilai SHAP ini nantinya yang akan memperhitungkan seberapa jauh masing-masing peubah berkontribusi untuk mengurangi atau menambah nilai dasar ke *output* yang dihasilkan.

Gambar 6 menampilkan plot nilai SHAP pada amatan data ke 100. Untuk kelas 1 (interaksi farmakodinamik), *probability output* yang dihasilkan adalah 1. Hal ini dilakukan karena dimulai dari *base value* 0,5, nilai ini bergeser ke kanan menuju 1 berdasarkan peubah-peubah yang berwarna merah. Peubah TPC = 0,682 merupakan peubah yang paling berkontribusi, kemudian CS = 0,105, dan TPCxCS = 0,072 yang mempengaruhi nilai SHAP. Gambar 7 menampilkan plot nilai SHAP secara menyeluruh untuk tiap amatan terhadap kelas 1.

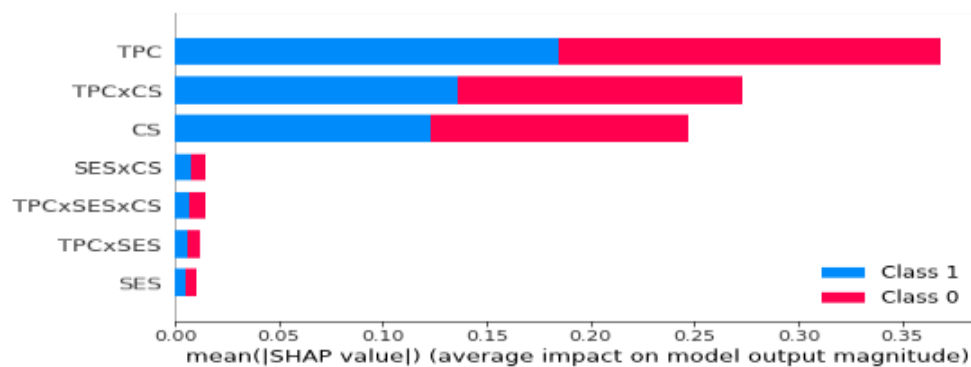


Gambar 6: Plot nilai SHAP pada amatan ke 100.



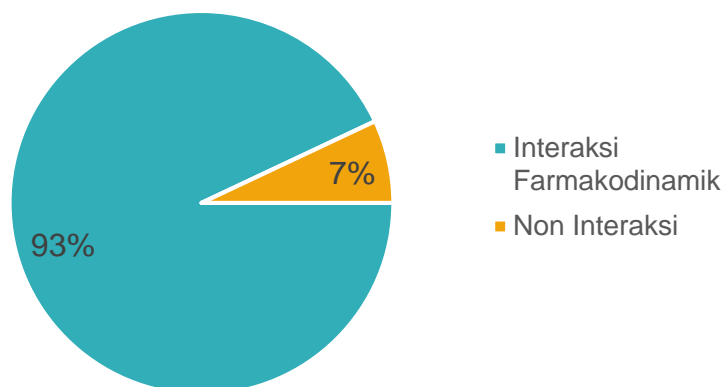
Gambar 7: Plot nilai SHAP pada seluruh amatan.

Kemudian untuk menggambarkan tingkat kepentingan peubah terhadap tiap amatan digunakan *Mean Absolute* (rata-rata absolut) dari nilai SHAP untuk setiap peubah dalam bentuk *stacked barplot* (Gambar 8). Plot ini menjelaskan, seberapa besar kontribusi suatu peubah terhadap prediksi tiap kelas. Peubah TPC, TPCxCS, dan CS sekali lagi menjadi peubah yang paling berkontribusi untuk membedakan kelas interaksi farmakodinamik pada model.



Gambar 8: *Mean absolute nilai SHAP* untuk setiap peubah di setiap kelas.

3.8 Hasil Klasifikasi Interaksi Senyawa



Gambar 9: Hasil klasifikasi interaksi senyawa.

Setelah model terbaik diperoleh, prediksi interaksi senyawa dilakukan terhadap 100 pasangan senyawa yang telah dihitung nilai TPC, SES, dan CS pada tahapan sebelumnya. Pada data yang ingin diprediksi dilakukan tahapan perlakuan yang sama seperti pembuatan data model yaitu membuat interaksi antar peubah menjadi TPC, SES, CS, TPC-SES, TPC-CS, SES-CS, dan TPC-SES-CS. Pada hasil prediksi diperoleh 93 pasangan senyawa berinteraksi farmakodinamik dan 7 senyawa tidak berinteraksi farmakodinamik. Tabel 6 menyajikan status interaksi senyawa aktif beserta informasi peubah penjelasnya.

Tabel 6: Status interaksi senyawa aktif.

Senyawa 1	Senyawa 2	TPC	SES	CS	Status
Benzoic acid	Melatonin	15,625	224,229	0,094	GSP/PD
Benzoic acid	Camphor	11,521	104,628	0,063	GSP/PD
Benzoic acid	Eucalyptol	15,622	127,414	0,058	GSP/PD
Benzoic acid	Glafenine	4,500	162,267	0,179	GSP/PD
Benzoic acid	Linalool	11,677	116,140	0,135	GSP/PD
Benzoic acid	Geraniol	10,318	114,500	0,185	GSP/PD
Benzoic acid	Curcumin	12,090	159,369	0,159	GSP/PD
Benzoic acid	Quercetin	12,319	128,149	0,182	GSP/PD
Benzoic acid	beta-Carotene	4,379	153,587	0,024	GSP/PD
Benzoic acid	Rutin	4,339	152,859	0,104	GSP/PD
Benzoic acid	Genistein	15,707	119,274	0,194	GSP/PD
Benzoic acid	Ascorbic acid	15,715	233,781	0,145	GSP/PD
Melatonin	Curcumin	13,837	211,889	0,186	GSP/PD
Propyl alcohol	Glafenine	1,674	141,638	0,190	GSP/PD
Camphor	Melatonin	11,942	233,865	0,099	GSP/PD
Camphor	Eucalyptol	11,926	71,385	0,366	GSP/PD
Camphor	Glafenine	3,669	160,277	0,091	GSP/PD
Camphor	Linalool	12,546	55,127	0,183	GSP/PD
Camphor	Geraniol	5,212	53,487	0,138	GSP/PD
Camphor	Curcumin	10,918	117,088	0,208	GSP/PD
Camphor	Quercetin	8,609	67,136	0,068	GSP/PD
Camphor	beta-Carotene	3,034	126,231	0,355	GSP/PD

Camphor	Rutin	4,029	131,517	0,124	GSP/PD
Camphor	Genistein	11,794	58,261	0,076	GSP/PD
Camphor	Ascorbic acid	12,204	258,140	0,117	GSP/PD
Eucalyptol	Melatonin	16,832	228,424	0,101	GSP/PD
Eucalyptol	Glafenine	4,585	154,836	0,173	GSP/PD
Eucalyptol	Linalool	12,447	62,112	0,365	GSP/PD
Eucalyptol	Geraniol	10,936	48,708	0,220	GSP/PD
Eucalyptol	Curcumin	13,137	120,943	0,143	GSP/PD
Eucalyptol	Quercetin	12,951	45,895	0,080	GSP/PD
Eucalyptol	beta-Carotene	5,798	135,557	0,392	GSN
Eucalyptol	Rutin	5,965	107,317	0,239	GSP/PD
Eucalyptol	Genistein	16,760	53,482	0,075	GSP/PD
Eucalyptol	Ascorbic acid	16,558	260,323	0,211	GSP/PD
Glafenine	Methylglyoxal	1,984	138,896	0,100	GSP/PD
Glafenine	Melatonin	4,824	202,065	0,126	GSP/PD
Glafenine	Linalool	3,780	158,307	0,155	GSP/PD
Glafenine	Lycopene	2,350	139,323	0,065	GSP/PD
Glafenine	Geraniol	4,800	144,903	0,161	GSP/PD
Glafenine	Squalene	1,242	142,430	0,066	GSP/PD
Glafenine	Curcumin	4,966	169,732	0,152	GSP/PD
Glafenine	Quercetin	6,171	142,089	0,150	GSP/PD
Glafenine	beta-Carotene	4,348	151,672	0,051	GSP/PD
Glafenine	Rutin	4,172	158,584	0,263	GSP/PD
Glafenine	Zeaxanthin	0,822	141,168	0,200	GSP/PD
Glafenine	Genistein	4,760	143,486	0,122	GSP/PD
Glafenine	Lutein	0,989	151,101	0,190	GSP/PD
Glafenine	Ascorbic acid	4,660	227,669	0,431	GSN
Linalool	Melatonin	12,469	245,255	0,104	GSP/PD
Linalool	Curcumin	11,436	120,833	0,241	GSP/PD
beta-Sitosterol	Glafenine	1,658	150,191	0,182	GSP/PD
Lycopene	Squalene	1,080	35,846	0,923	GSP/PD
Geraniol	Melatonin	10,765	244,052	0,123	GSP/PD
Geraniol	Linalool	5,731	18,653	0,600	GSP/PD
Geraniol	Curcumin	7,072	107,429	0,295	GSN
Chlorogenic acid	Benzoic acid	7,076	119,681	0,090	GSP/PD
Chlorogenic acid	Melatonin	8,097	237,033	0,076	GSP/PD
Chlorogenic acid	Camphor	6,937	58,668	0,218	GSP/PD
Chlorogenic acid	Eucalyptol	7,557	53,890	0,302	GSP/PD
Chlorogenic acid	Glafenine	4,357	150,084	0,252	GSP/PD
Chlorogenic acid	Linalool	7,181	23,834	0,214	GSP/PD
Chlorogenic acid	Geraniol	0,479	10,430	0,139	GSP/PD
Chlorogenic acid	Curcumin	7,949	112,611	0,252	GSP/PD
Chlorogenic acid	Quercetin	6,214	24,080	0,205	GSP/PD
Chlorogenic acid	beta-Carotene	0,045	91,207	0,168	GSP/PD
Chlorogenic acid	Rutin	1,068	114,644	0,355	GSP/PD
Chlorogenic acid	Genistein	9,385	15,204	0,125	GSP/PD
Chlorogenic acid	Ascorbic acid	7,909	273,071	0,460	GSN
Quercetin	Melatonin	14,154	235,108	0,135	GSP/PD

Quercetin	Linalool	9,170	32,302	0,153	GSP/PD
Quercetin	Geraniol	10,104	18,898	0,184	GSP/PD
Quercetin	Curcumin	13,667	104,616	0,351	GSP/PD
Quercetin	beta-Carotene	6,001	105,747	0,063	GSP/PD
Quercetin	Rutin	7,177	106,649	0,542	GSN
Quercetin	Genistein	17,805	23,672	0,605	GSN
Quercetin	Ascorbic acid	12,785	271,147	0,203	GSP/PD
beta-Carotene	Melatonin	5,937	224,855	0,072	GSP/PD
beta-Carotene	Linalool	3,598	105,501	0,349	GSP/PD
beta-Carotene	Geraniol	4,542	92,097	0,297	GSP/PD
beta-Carotene	Curcumin	6,467	146,379	0,175	GSP/PD
beta-Carotene	Rutin	4,817	136,925	0,110	GSP/PD
beta-Carotene	Genistein	7,181	96,871	0,073	GSP/PD
beta-Carotene	Ascorbic acid	5,330	243,530	0,097	GSP/PD
Rutin	Melatonin	6,450	215,623	0,114	GSP/PD
Rutin	Linalool	4,631	122,866	0,202	GSP/PD
Rutin	Geraniol	4,598	109,463	0,182	GSP/PD
Rutin	Curcumin	8,239	158,133	0,294	GSP/PD
Rutin	Genistein	9,127	114,237	0,342	GSP/PD
Rutin	Ascorbic acid	5,655	239,032	0,341	GSN
Zeaxanthin	Lutein	0,002	42,632	0,915	GSP/PD
Genistein	Melatonin	17,879	237,300	0,136	GSP/PD
Genistein	Linalool	12,338	23,427	0,157	GSP/PD
Genistein	Geraniol	10,622	10,023	0,195	GSP/PD
Genistein	Curcumin	16,697	112,203	0,283	GSP/PD
Genistein	Ascorbic acid	16,538	273,339	0,194	GSP/PD
Ascorbic acid	Melatonin	16,686	178,972	0,105	GSP/PD
Ascorbic acid	Linalool	12,564	269,530	0,254	GSP/PD
Ascorbic acid	Geraniol	10,563	274,842	0,255	GSP/PD
Ascorbic acid	Curcumin	13,113	243,209	0,222	GSP/PD

Berdasarkan penelusuran yang mengidentifikasi efek sinergitas antar pasangan senyawa aktif tanaman jamu Diabetes Melitus Tipe 2 yang dilakukan oleh Syahrir *et al.* (2016) menggunakan metode *Network target-based Identification of Multicomponent Synergy* (NIMS), dari 93 pasangan senyawa yang berinteraksi farmakodinamik, terdapat tiga pasang senyawa yang secara signifikan berinteraksi sinergi yaitu *Eucalyptol-Linalool*, *Eucalyptol-Geraniol*, dan *Linalool-Geraniol*.

4. Simpulan

Berdasarkan hasil dan pembahasan pada penelitian ini, dapat disimpulkan bahwa terdapat 100 interaksi senyawa yang diperoleh dari tanaman jamu untuk pengobatan Diabetes Melitus Tipe 2 yaitu pare, sembung, bratawali, dan jahe. Dengan menggunakan model klasifikasi terbaik yaitu *Random Forest* dengan peubah penjelas TPC, SES, dan CS, 93 interaksi senyawa diidentifikasi berinteraksi farmakodinamik dan hanya 7 pasangan senyawa yang tidak berinteraksi farmakodinamik. Pasangan senyawa yang tidak berinteraksi yaitu *Eucalyptol-beta-Carotene*, *Glaflenine-Ascorbic acid*, *Geraniol-Curcumin*, *Chlorogenic acid-Ascorbic acid*, *Quercetin-Rutin*, *Quercetin-Genistein*, dan *Rutin-Ascorbic acid*.

Daftar Pustaka

- Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., ... others. (2012). KNAPSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant and Cell Physiology*, 53(2): e1–e1.
- Altaf-Ul-Amin, M., Tsuji, H., Kurokawa, K., Ashahi, H., Shinbo, Y., & Kanaya, S. (2007). A density-periphery based graph clustering software developed for detection of protein complexes in interaction networks. *2007 International Conference on Information and Communication Technology*, 37–42. IEEE.
- Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J., & Bork, P. (2008). Drug target identification using side-effect similarity. *Science*, 321(5886): 263–266.
- Chapman, J. (2008). *Chapman and Hall dictionary of natural products*. CRC Press, Hampden Data Services Ltd.
- Dong, J., Cao, D.-S., Miao, H.-Y., Liu, S., Deng, B.-C., Yun, Y.-H., ... Chen, A. F. (2015). ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *Journal of Cheminformatics*, 7(1): 1–10.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8): 861–874.
- Hasnita, H., Afendi, F. M., & Fitrianto, A. (2020). Perbandingan beberapa metode klasifikasi dalam memprediksi interaksi farmakodinamik. *Indonesian Journal of Statistics and Its Applications*, 4(1): 11–21.
- Huang, J., Niu, C., Green, C. D., Yang, L., Mei, H., & Han, J.-D. J. (2013). Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. *PLoS Computational Biology*, 9(3): e1002998.
- Katno, P. S., Prapti, I., Rahmawati, N., & Mujahid, R. (2008). Tingkat Manfaat, Keamanan dan Efektifitas Tanaman Obat dan Obat Tradisional. Balai Penelit Tanam Obat Tawangmangu.
- Merle, L., Laroche, M.-L., Dantoine, T., & Charmes, J.-P. (2005). Predicting and preventing adverse drug reactions in the very old. *Drugs & Aging*, 22(5): 375–392.
- Mulia, I. (2017). Model Prediksi Interaksi Senyawa Protein Menggunakan Fungsi Kemiripan dan *Fingerprint Klekota-Roth* (PhD Thesis). Bogor Agricultural University (IPB).
- Nurishmaya, M. R. (2014). Pendekatan Bioinformatika Formulasi Jamu Baru Berkhasiat Antidiabetes dengan Ikan Zebra (*Danio rerio*) sebagai Hewan Model.
- Syahrir, N. H. A., Afendi, F. M., & Susetyo, B. (2016). Efek sinergis bahan aktif tanaman obat berbasis jejaring dengan protein target. *Jurnal Jamu Indonesia*, 1(1): 35–46.
- Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research*, 37(suppl_2): W623–W633.
- Winata, H. M., Afendi, F. M., & Fitrianto, A. (2019). Peningkatan akurasi klasifikasi interaksi farmakodinamik obat berbasis seleksi pasangan obat takberinteraksi. *Indonesian Journal of Statistics and Its Applications*, 3(3): 247–259.