

## Target prediction of compounds on jamu formula using nearest profile method

### Prediksi Target Dari Senyawa Pada Formula Jamu Dengan Metode Profil Terdekat

Nur Hilal A. Syahrir<sup>1\*</sup>, Sumarheni<sup>2</sup>, Supri Bin Hj Amir<sup>3\*</sup>, Hedi Kuswanto<sup>4</sup>

#### Abstract

Jamu is one of Indonesia's cultural heritage, which consists of several plants that have been practiced for centuries in Indonesian society to maintain health and treat diseases. One of the scientification efforts of Jamu to reveal its mechanism is to predict the target-protein of the active ingredients of the Jamu. In this study, the prediction of the target compound for Jamu was carried out using a supervised learning approach involving conventional medicinal compounds as training data. The method used in this study is the closest profile method adopted from the nearest neighbor algorithm. This method is implemented in drug compound data to construct a learning model. The AUC value for measuring performance of the three implemented models is 0.62 for the fixed compound model, 0.78 for the fixed target model, and 0.83 for the mixed model. The fixed compound model is then used to construct a prediction model on the herbal medicine data with an optimal threshold value of 0.91. The model produced 10 potential compounds in the herbal formula and its 44 unique protein targets. Even though it has many limitations in obtaining a good performance, the closest profile method can be used to predict the target of the herbal compound whose target is not yet known.

**Keywords:** supervised learning, jamu, compound-target protein

#### Abstrak

Jamu merupakan salah satu warisan budaya Indonesia yang terdiri dari beberapa tanaman yang telah dipraktikkan selama berabad-abad di masyarakat Indonesia untuk menjaga kesehatan dan mengobati penyakit. Salah satu upaya saintifikasi jamu untuk mengungkap mekanisme kerja jamu adalah prediksi target-protein dari bahan aktif jamu. Pada penelitian ini, prediksi target senyawa jamu dilakukan dengan pendekatan supervised learning dengan melibatkan senyawa obat konvensional sebagai data latih. Metode yang digunakan pada

<sup>\*</sup>Program Studi Ilmu Komputer Universitas Hasanuddin, <sup>2</sup>Program Studi Farmasi Universitas Hasanuddin, <sup>4</sup>Program Studi Statistika Universitas Hasanuddin.

**Email:** <sup>1</sup>nurhilal@unhas.ac.id, <sup>2</sup>sumarheni@gmail.com, <sup>3</sup>supriamir@unhas.ac.id,

<sup>4</sup>hedikuswanto454@gmail.com



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

**Nur Hilal A. Syahrir, Sumarheni, Supri Bin Hj Amir, Hedi Kuswanto**  
*Jurnal Matematika, Statistika & Komputasi*

penelitian ini adalah metode profil terdekat yang diadopsi dari algoritma *nearest neighbour*. Metode ini diimplementasikan pada data senyawa obat untuk mengkonstruksi model learning. Nilai AUC untuk mengukur performa dari ketiga model yang diimplementasikan adalah 0.62 untuk *model fixed compound*, 0.78 untuk *model fixed target*, dan 0.83 untuk *mixed model*. Model *fixed compound* lalu digunakan untuk mengkonstruksi model prediksi pada data jamu dengan nilai ambang batas optimal adalah 0.91. Model tersebut menghasilkan 10 senyawa potensial pada formula jamu dan 44 target protein unik yang ditargetkan oleh satu senyawa. Meskipun memiliki banyak keterbatasan dalam memperoleh performa yang baik, metode profil terdekat dapat digunakan untuk memprediksi target dari senyawa jamu yang belum diketahui targetnya.

**Kata kunci:** supervised learning, jamu, senyawa-target protein

## 1. PENDAHULUAN

Jamu merupakan salah satu warisan budaya Indonesia yang terdiri dari beberapa tumbuhan yang telah dipraktikkan selama berabad-abad di masyarakat Indonesia untuk menjaga kesehatan dan mengobati penyakit [5]. Meskipun penggunaan obat-obatan modern (konvensional) di Indonesia makin meningkat secara signifikan, namun penggunaan jamu sebagai salah satu upaya pengobatan alternatif juga cukup populer digunakan oleh masyarakat Indonesia. Hal tersebut didukung oleh keterjangkauan dan ketersediaan bahan jamu di Indonesia. Upaya pengembangan jamu mulai digalakkan dengan adanya Peraturan Menteri Kesehatan RI No.002/Menkes/Per/I/2010 untuk melakukan terobosan dengan program saintifikasi Jamu [1]. Program ini berupaya memberi landasan ilmiah (*evidence-based*) bagi jamu yang sebelumnya lebih berdasarkan landasan pengalaman (*empirical-based*).

Salah satu upaya saintifikasi jamu untuk mengungkap mekanisme kerja jamu adalah prediksi target-protein dari bahan aktif jamu. Upaya prediksi target senyawa jamu dapat dilakukan di laboratorium basah dengan mengekstrak senyawa dan memadukannya dengan protein target tertentu. Upaya ini tentu membutuhkan waktu, biaya, dan tenaga yang relatif besar jika jumlah senyawa yang akan diujikan di laboratorium juga besar. Oleh karena itu, pada penelitian ini, upaya prediksi target protein dari senyawa jamu dilakukan dengan pendekatan komputasi untuk menghasilkan prediksi target protein dari senyawa dalam kurun waktu yang lebih singkat pada jumlah senyawa yang cukup besar.

Berbagai metode komputasi telah dikembangkan untuk menganalisis dan memprediksi interaksi senyawa-protein, seperti simulasi *molecular-docking* [7] dan metode berbasis *network* [12]. Namun, kedua teknik tersebut memiliki keterbatasan untuk digunakan dalam memprediksi target senyawa dari jamu. Misalnya, simulasi *docking* tidak dapat diterapkan pada protein yang struktur 3D-nya tidak diketahui. Sementara struktur 3 dimensi pada senyawa yang terdapat di jamu masih sulit diidentifikasi pada pangkalan data. Pendekatan dengan berbasis *network* juga masih terbatas dilakukan pada prediksi target jamu dikarenakan tidak tersedianya histori *network* dari senyawa jamu.

Pada penelitian ini, prediksi target senyawa jamu dilakukan dengan pendekatan *supervised learning* dengan melibatkan senyawa obat konvensional sebagai data latih. Metode yang digunakan pada penelitian ini adalah metode profil terdekat yang diadopsi dari algoritma *nearest neighbour*. Metode ini diimplementasikan pada data senyawa obat untuk mengkonstruksi model learning. Setelah model learning diperoleh, model tersebut diimplementasikan untuk memperoleh prediksi target dari senyawa pada formula jamu. Formula jamu yang akan diprediksi targetnya terdiri dari empat tanaman obat, yaitu *Tinospora crispa*, *Zingiber officinale*, *Momordica*

*charantia*, dan *Blumea balsamifera*. Hasil dari metode komputasi ini bergantung pada ketersediaan data. Jumlah data yang berbeda dapat menghasilkan prediksi yang berbeda pula.

## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari berbagai pangkalan data non-komersil. Data senyawa yang digunakan pada penelitian ini terdiri atas daftar senyawa obat yang telah diuji secara klinis dan daftar senyawa yang terkandung pada formula jamu yang terdiri dari empat tanaman yaitu tanaman Bratawali (*Tinospora crispa*), Jahe (*Zingiber officinale*), Pare (*Momordica charantia*), dan Sembung (*Blumea balsamifera*). Daftar senyawa diperoleh pada pangkalan data DrugBank (<https://www.drugbank.ca/>) [11]. Kategori daftar senyawa obat yang digunakan adalah kategori FDA (*Food and Drug Administration*) *Approved*. Senyawa obat yang telah ditarik dari pasaran ataupun yang masih dalam tahapan ekperimental tidak digunakan pada penelitian ini. Terdapat 2140 senyawa obat yang tergolong dalam kategori *FDA-Approved*, namun hanya 1810 senyawa obat yang dapat diidentifikasi struktur molekulnya dan direpresantesikan dalam bentuk dua dimensi. Selain data senyawa, 2408 target protein dari senyawa obat dan data mengenai interaksi antar senyawa dan target protein dikumpulkan melalui pangkalan data *DrugBank* [5].

Data senyawa jamu diperoleh dari pangkalan data *KnapsackFamily* [2] dan *The Dictionary of Natural Products* [9]. Terdapat 595 data senyawa yang tercatat pada empat tanaman tersebut, namun hanya 103 senyawa yang dapat diidentifikasi struktur molekulnya pada pangkalan data *PubChem* [10]. Senyawa pada tanaman terdiri dari 3 senyawa pada tanaman Bratawali (*Tinospora crispa*), 61 senyawa pada tanaman Jahe (*Zingiber officinale*), 28 senyawa pada tanaman Pare (*Momordica charantia*), dan 11 senyawa pada tanaman Sembung (*Blumea balsamifera*).

### 2.2 Metode Analisis

Pendekatan komputasi yang digunakan pada penelitian ini merupakan pendekatan berbasis *supervised-learning* dimana interaksi dari senyawa dan target merupakan label dari data. Data terdiri atas data latih, data validasi, dan data testing. Data latih dan data validasi dilakukan pada data senyawa obat, sementara data testing dilakukan pada senyawa jamu. Tahapan yang dilakukan pada penelitian teridiri atas tiga tahapan utama:

1. Tahapan konstruksi atribut dan label learning.
  - a. Menkonstruksi graf bipartit pada data interaksi senyawa obat dan target sebagai label dari algoritma *supervised-learning*.
  - b. Menghitung nilai kemiripan senyawa obat kimiawi antar dua senyawa obat dalam bentuk matriks kemiripan struktur senyawa sebagai fitur/atribut.
  - c. Menghitung nilai kemiripan protein target dalam bentuk matriks kemiripan protein sebagai fitur.
2. Tahapan konstruksi model supervised learning menggunakan metode profil terdekat
3. Tahapan aplikasi model learning pada senyawa jamu.
  - a. Menghitung kemiripan senyawa jamu dan senyawa obat
  - b. Mengimplementasikan model profil terdekat pada senyawa jamu

### 2.2.1 Kemiripan Struktur Senyawa.

Kemiripan struktur kimiawi senyawa dihitung berdasarkan asumsi bahwa senyawa dengan struktur kimia yang mirip biasanya berikatan secara fungsional protein target terkait. Kemiripan struktur kimiawi antar dua pasangan senyawa dapat dihitung dengan menggunakan koefisien Tanimoto [7]. Kemiripan struktur kimiawi antara dua molekul A dan B dapat dinotasikan sebagai:

$$S_{c(A,B)} = \frac{c}{[a + b - c]} \quad (2.1)$$

dimana  $a$  adalah jumlah bits pada molekul A,  $b$  adalah jumlah bits pada molekul B, sedangkan  $c$  adalah jumlah bit yang ada di kedua molekul. Semua SMILES senyawa obat dan jamu diubah menjadi bilangan biner menggunakan sidik jari ECFPs (*Extended Connectivity Fingerprint*).

Sidik jari tersebut merupakan koleksi 1024-bit unik sebagai representasi substruktur senyawa kimia yang dihasilkan dengan memecah setiap gabungan SMILES.

### 2.2.2 Kemiripan protein target

Protein target direpresentasikan sebagai urutan asam amino dalam bentuk string/karakter. Kemiripan antar pasangan protein dihitung menggunakan Algoritma Smith-Waterman [8]. Konsep dasar algoritma tersebut adalah melakukan *alignment* urutan asam amino secara lokal, dengan menentukan daerah yang mirip antara dua string urutan asam amino. Daerah yang mirip diperoleh dengan membandingkan segmen dari semua kemungkinan panjang untuk mengoptimalkan ukuran kemiripan.

Ide *alignment* dua urutan asam amino pada algoritma Smith-Waterman terletak pada pengisian matriks kemiripan  $H$  yang berukuran  $n \times m$ , dimana  $n$  adalah jumlah elemen dalam urutan kueri asam amino A dan  $m$  adalah jumlah elemen dalam urutan referensi asam amino B. Setiap nilai  $H[i, j]$  adalah ukuran kemiripan dua urutan asam amino pada urutan kueri hingga elemen ke- $i$  dan urutan referensi hingga elemen ke- $j$ . Nilai-nilai matriks  $H[i, j]$  dihitung menggunakan pemrograman dinamis sesuai dengan rumus:

Inisialisasi:

$$\begin{aligned} H[i, 0] &= 0, 0 \leq i \leq n \\ H[0, j] &= 0, 0 \leq j \leq m \end{aligned} \quad (2.2)$$

Rekursi:

$$H[i, j] = \max \begin{cases} 0 \\ H[i-1, j] + wk \\ H[i, j-1] + wk \\ H[i-1, j-1] + S[i, j] \end{cases}, 1 \leq i \leq n, 1 \leq j \leq m \quad (2.3)$$

dimana  $S[i, j]$  adalah *match / mismatch* skor yang telah ditentukan dan  $wk$  adalah konstanta *gap penalty*. Skor Smith-Waterman kemudian ditentukan melalui tracing pada matriks BLOSUM.

Matriks kemiripan protein target kemudian diperoleh dengan melakukan normalisasi pada skor Smith-Waterman. Normalisasi pasangan protein ( $P_A, P_B$ ) dihitung dengan formula:

$$S_g(P_A, P_B) = \frac{SW(P_A, P_B)}{\sqrt{SW(P_A, P_A)} \times \sqrt{SW(P_A, P_B)}} \quad (2.4)$$

### 2.2.3 Metode profil terdekat

Untuk memprediksi interaksi baru antara senyawa dan target, metode profil terdekat dapat dilakukan. Konsep dasar *supervised learning* dengan metode profil terdekat ialah interaksi senyawa-protein merupakan representasi dari sebuah graf bipartite yang mampu memprediksi apakah sebuah senyawa C menargetkan *Protein of Interest* (POI) atau tidak, dan apakah sebuah protein P ditargetkan oleh *Compound of Interest* atau tidak (COI) [4].

Tahapan prediksi menggunakan algoritma supervised learning pada umumnya menggunakan dua tahapan, pertama dilakukan dengan membuat pembelajaran pada data latih/training yang melibatkan data label sebagai “*gold standard*”. Pada proses kedua, model pembelajaran diaplikasikan pada senyawa dan protein dengan menghilangkan “*gold standard*” untuk membuat inferensi mengenai interaksi dari senyawa dan protein.

Metode profil terdekat merupakan salah satu metode naïve yang dikembangkan berdasarkan ide dasar dari Metode *Nearest Neighbour* [13]. Misalkan kita mempunyai himpunan senyawa  $\{c_i\}_{i=1}^{n_c}$  dan protein target yang diketahui  $\{g_j\}_{j=1}^{n_g}$ , dimana  $n_c$  merupakan jumlah senyawa dan  $n_g$  merupakan jumlah protein. Diketahui pula pola interaksi dari senyawa  $c_i$  terhadap protein dan interaksi dari protein  $g_j$  terhadap senyawa. Sehingga diperoleh interaksi profile  $\mathbf{x}_{c_i}$  yang didefinisikan sebagai string 1 dan 0 (vektor yang berukuran  $n_g$ ), dimana 1 merepresentasikan adanya interaksi antar senyawa dan target dan 0 merepresentasikan tidak adanya interaksi antar senyawa target pada himpunan protein  $g_j$  ( $j = 1, 2, \dots, n_g$ ). Sebaliknya diketahui pula interaksi profile  $\mathbf{y}_{g_j}$  senyawa dan interaksi terhadap senyawa  $c_i$ . Sehingga diperoleh interaksi profile  $\mathbf{y}_{g_j}$  yang didefinisikan sebagai string 1 dan 0 (vector yang berukuran  $n_c$ ), dimana 1 merepresentasikan adanya interaksi antar target dan senyawa dan 0 merepresentasikan tidak adanya interaksi antar target dan senyawa pada himpunan protein ( $j = c_i, i = 1, 2, \dots, n_c$ ).

Misalkan terdapat himpunan dari profil interaksi  $\{\mathbf{x}_{c_i}\}_{i=1}^{n_c}$  dan  $\{\mathbf{y}_{g_j}\}_{j=1}^{n_g}$ , maka prediksi senyawa baru  $\mathbf{x}_{c_{new}}$  dan target protein baru  $\mathbf{y}_{g_{new}}$  dapat diperoleh. Prediksi senyawa baru dapat diperoleh dengan formula:

$$\mathbf{x}_{c_{new}} = S_c(c_{new}, c_{nearest}) \mathbf{x}_{c_{nearest}} \quad (2.5)$$

dimana  $S_c$  merupakan kemiripan struktur senyawa kimia dan  $c_{nearest}$  merupakan senyawa yang paling mirip dengan senyawa  $c_{new}$ . Sementara prediksi target baru dari suatu senyawa dapat diperoleh dengan formula:

$$\mathbf{y}_{g_{new}} = S_g(g_{new}, g_{nearest}) \mathbf{y}_{g_{nearest}} \quad (2.6)$$

dimana  $S_g$  merupakan kemiripan sekuens protein dan  $g_{nearest}$  merupakan target protein yang paling mirip dengan target protein  $g_{new}$ .

### 2.2.4 Evaluasi kinerja prediksi

Performa prediksi target dari senyawa pada jamu menggunakan metode profil terdekat dievaluasi menggunakan *Leave-One-Out-Cross-Validation* (LOOCV). Pada prinsipnya metode ini merupakan kasus khusus dari *Cross Validation* (CV) dimana jumlah *fold* sama dengan jumlah pengamatan ( $n$ ) pada dataset. Dengan demikian, algoritme pembelajaran diterapkan sekali untuk setiap pengamatan, menggunakan semua pengamatan lainnya sebagai data latih dan

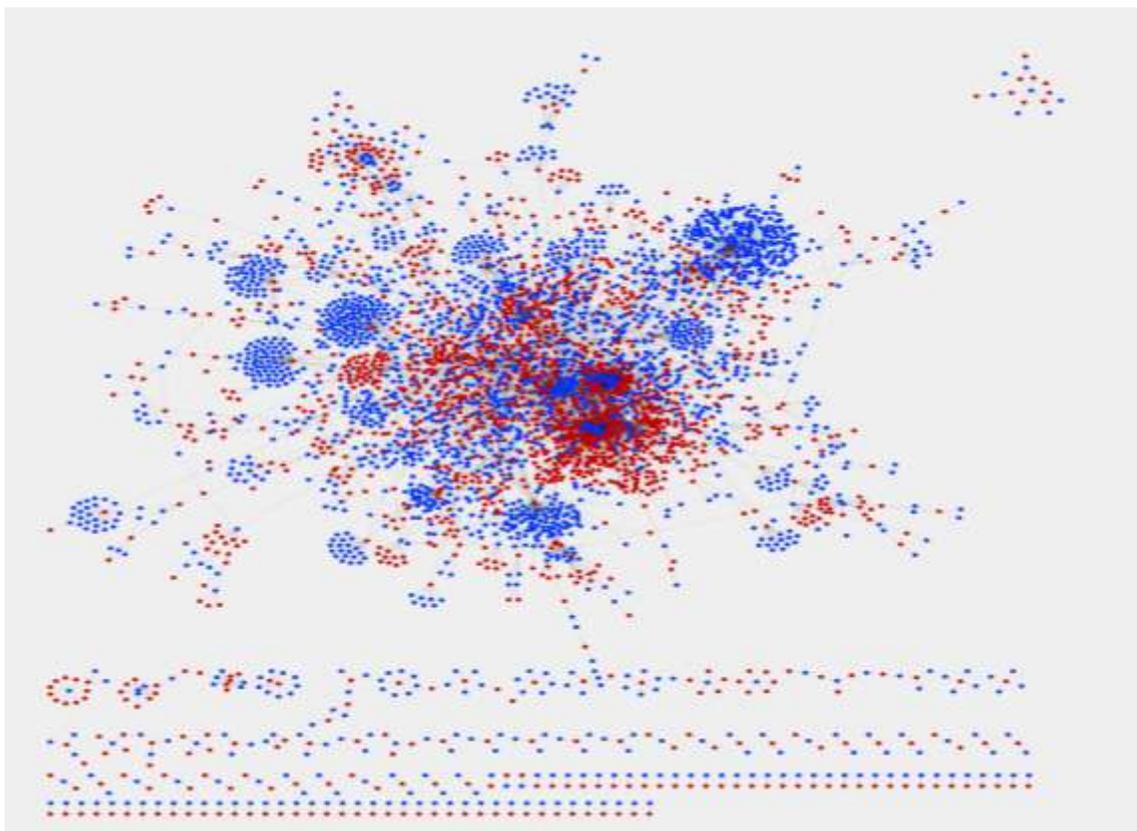
menggunakan pengamatan yang dipilih sebagai data uji dengan item tunggal. Proses ini hampir serupa dengan metode estimasi Jacknife pada metode statistika.

Ukuran estimasi kinerja pada penelitian ini, menggunakan ukuran estimasi kinerja pada pembelajaran *supervised*, seperti akurasi, sensitivitas (*Recall*), spesifitas, presisi, dan Skor F1 yang diperoleh dari matriks klasifikasi (*confusion matrix*) [6].

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Konstruksi graf interaksi senyawa-target

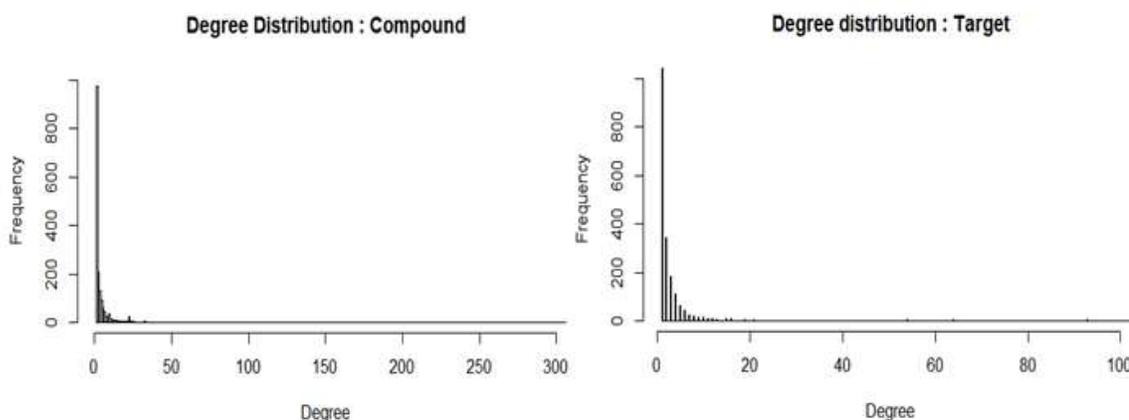
Pada penelitian ini konstruksi jejaring interaksi senyawa-target dilakukan dengan menggunakan representasi graf bipartit (Gambar 1). Simpul (*node*) graf terdiri dari dua himpunan terpisah, yaitu himpunan senyawa obat dan himpunan protein target, sementara sisi (*edge*) merupakan interaksi antar kedua himpunan senyawa dan target. Interaksi tersebut menunjukkan pasangan senyawa dan target diketahui memiliki hubungan yang telah melalui tahapan uji klinis.



**Gambar 1.** Graf interaksi senyawa-target obat pada data latih. Simpul berwarna merah merupakan entitas senyawa dan simpul berwarna biru merupakan entitas target.

Gambar 2 menunjukkan distribusi derajat (*degree distribution*) senyawa obat dan target protein pada jejaring interaksi senyawa-target. Derajat dari simpul senyawa merupakan jumlah target yang mempunyai hubungan dengan senyawa. Sementara derajat dari simpul target merupakan jumlah senyawa yang mempunyai hubungan dengan target. Distribusi derajat pada senyawa maupun target mengindikasikan adanya karakteristik *power law distribution* yang berarti

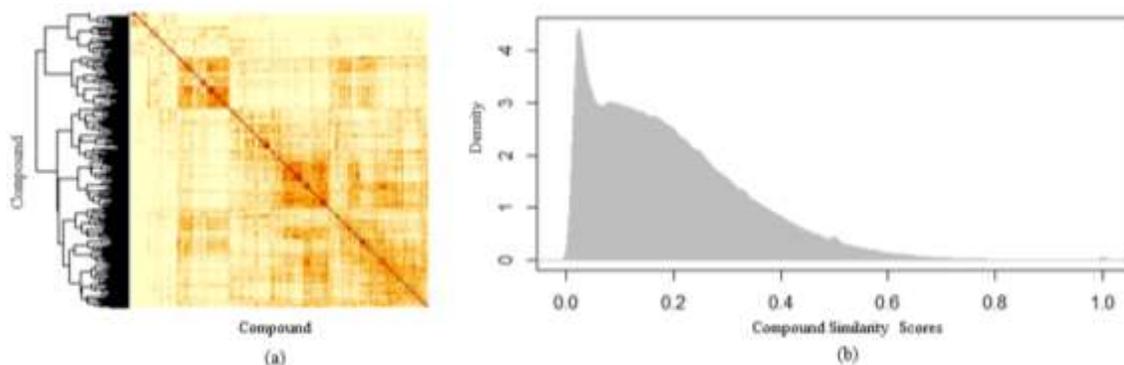
bahwa hanya sedikit target protein yang merupakan target dari jumlah senyawa yang banyak dan banyak target protein yang menjadi target dari jumlah senyawa yang sedikit. Karakteristik yang sama berlaku pada derajat distribusi target, bahwa sedikit senyawa yang menargetkan banyak protein dan banyak senyawa yang menargetkan sedikit protein.



**Gambar 2.** Sebaran derajat dari senyawa dan target

### 3.2 Kemiripan struktur kimiawi senyawa obat

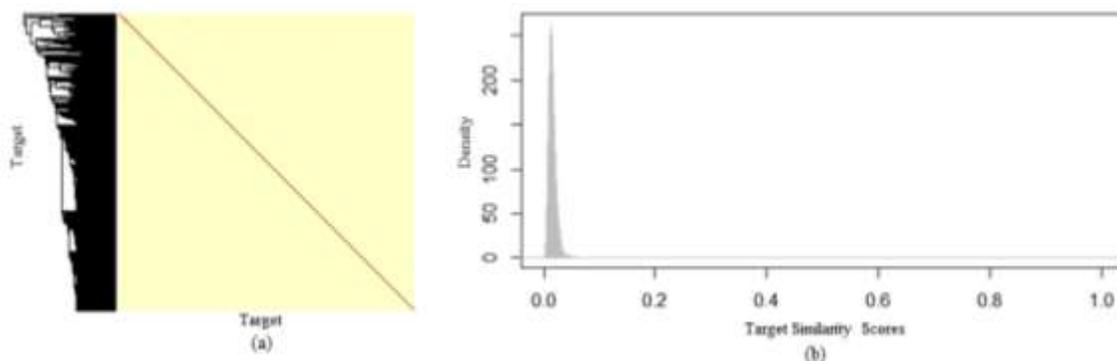
Kemiripan struktur kimiawi pada senyawa mengacu kepada kemiripan elemen atau unsur senyawa [3]. Kuantifikasi kemiripan senyawa pada pasangan senyawa dilakukan dengan asumsi bahwa senyawa yang mirip secara struktur akan cenderung menargetkan protein yang serupa [13]. Kuantifikasi kemiripan struktur pada senyawa obat merupakan atribut learning pada model supervised learning. Skor kemiripan senyawa berada pada range nilai 0 hingga 1. Semakin mirip suatu pasangan senyawa maka skor kemiripan senyawa akan mendekati 1. Begitu pula sebaliknya, semakin berbeda suatu pasangan senyawa maka skor kemiripan akan mendekati nol. Gambar (a) berikut menunjukkan skor semua pasangan senyawa pada senyawa obat. Semakin terang, maka semakin tinggi kemiripan antar pasangan senyawa, dan sebaliknya. Kemiripan antar pasangan senyawa pada Gambar 3(a) menunjukkan bahwa sebagian besar pasangan senyawa cenderung tidak memiliki kemiripan tinggi. Hal tersebut juga dapat mengindikasikan bahwa kemungkinan pasangan senyawa menargetkan protein target yang berbeda. Selain itu, pada Gambar 3(b) dapat dilihat bahwa proporsi senyawa dengan kemiripan lebih besar dari 0.4 sangat kecil.



**Gambar 3.** (a) Heat map skor kemiripan struktur kimiawi antar pasangan senyawa obat. (b). Plot densitas skor kemiripan senyawa.

### 3.3 Kemiripan urutan asam amino target protein

Model supervised learning dengan metode profil terdekat tidak hanya mampu memprediksi apakah sebuah senyawa *C* menargetkan *Protein of Interest* (POI) namun juga dapat memprediksi apakah sebuah protein *P* ditargetkan oleh *Compound of Interest* (COI). Hal tersebut didasarkan oleh asumsi bahwa pasangan protein target yang memiliki kemiripan sekuens asam amino yang tinggi cenderung berinteraksi dengan obat yang serupa dan karena protein tersebut dekat dalam network [13]. Berbeda dengan kemiripan struktur kimiawi pada pasangan obat, kuantifikasi skor kemiripan sekuens asam amino dari pasangan target memiliki range nilai yang relatif kecil (Gambar 4). Skor kemiripan antar pasangan protein hanya berada di sekitar 0.0001 hingga kurang dari 0.01. Dengan demikian, dapat diketahui bahwa kecenderungan pasangan protein menargetkan senyawa yang berbeda.



**Gambar 4.** (a) Skor kemiripan antar pasangan target protein. (b) Plot densitas skor kemiripan target protein

### 3.4 Metode profil terdekat

Metode profil terdekat pada senyawa obat diimplementasikan pada tiga jenis skenario. Skenario pertama, dilakukan prediksi target senyawa dengan kemiripan antar senyawa sebagai atribut. Pada skenario kedua dilakukan prediksi senyawa dari suatu target dengan kemiripan antar protein sebagai atribut. Terakhir, prediksi senyawa dilakukan dengan menggabungkan skenario pertama dan skenario ketiga.

Tabel 1. Performa metode profil terdekat pada tiga model.

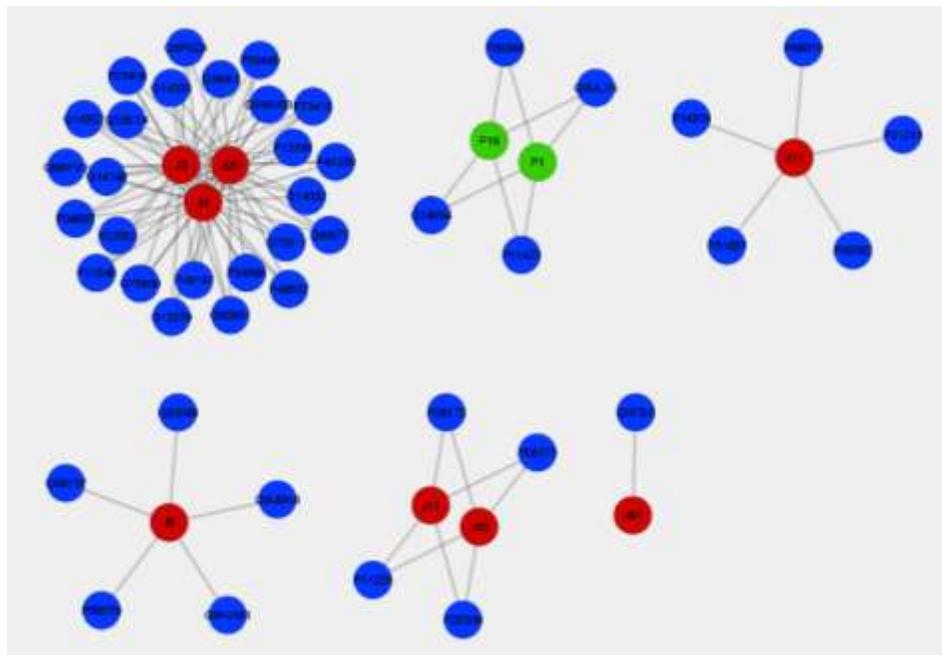
	Model 1 (Fixed Compound)	Model 2 (Fixed Target)	Model 3 (Mixed Model)
Ambang Batas Optimal	0.910	0.320	0.930
AUC	0.623	0.780	<b>0.832</b>
Sensitifitas / Recall	0.120	<b>0.491</b>	0.120
Spesifitas	0.999	0.999	0.999
Presisi	0.316	<b>0.657</b>	0.315
Akurasi	0.997	0.998	0.997
Skor F1	0.173	<b>0.562</b>	0.174

Tabel di atas menunjukkan bahwa penggabungan model 1 dan model 2 pada model 3 menunjukkan penambahan nilai AUC pada kurva ROC. Hal tersebut menunjukkan adanya peningkatan performa dengan penggabungan skor kemiripan senyawa dan skor kemiripan protein. Akurasi dari ketiga model juga cukup baik. Namun, meskipun ketiga model memiliki nilai akurasi yang baik, salah satu kekurangan pada penelitian ini adalah rendahnya nilai presisi dan sensitifitas. Hal ini disebabkan adanya ketidakseimbangan pada label/target. Pada penelitian ini, label pada data senyawa maupun data target protein terdiri dari beberapa label (*multi-label*) untuk setiap instances dari data, sehingga teknik menangani data tidak seimbang dengan *oversampling* ataupun *undersampling* menjadi sulit untuk dilakukan. Hal tersebut mengimplikasikan kurang baiknya kemampuan model untuk memprediksi kelas minoritas.

### 3.5 Implementasi metode profil terdekat untuk prediksi target jamu

Metode profil terdekat pada latih kemudian diimplementasikan untuk memprediksi target protein dari senyawa yang terdapat pada formula jamu. Karena fokus penelitian ini untuk memperoleh target dari protein dari senyawa jamu, *model fixed compound* lalu digunakan untuk mengkonstruksi model prediksi pada data jamu. Nilai ambang batas optimal adalah 0.91 yang merupakan ambang batas optimal dari *model fixed compound*. Gambar 5 merupakan *Network* prediksi target dari senyawa formula jamu. Node hijau dan merah merupakan senyawa dari formula jamu. Node hijau merupakan senyawa dari tanaman pare (*Momordica charantia*) dan node merah merupakan senyawa dari tanaman jahe (*Zingiber officinale*). Sementara node biru merupakan protein target dari senyawa.

Pada jejaring tersebut dapat dilihat bahwa model profil terdekat menghasilkan prediksi dari dua tanaman saja yaitu pare dan jahe. Hal tersebut juga merupakan implikasi dari terbatasnya informasi senyawa lainnya yaitu Bratawali dan Sembung pada pangkalan data KnapsackFamily [2] dan *The Dictionary of Natural Products* [9], yaitu terdiri dari 3 senyawa pada tanaman Bratawali (*Tinospora crispa*) dan 11 senyawa pada tanaman Sembung (*Blumea balsamivera*). Prediksi target jamu dengan metode profil terdekat dengan ambang batas 0.91 menghasilkan 10 senyawa potensial pada formula jamu dengan 44 target protein unik dengan 25 protein target ditargetkan oleh tiga senyawa, 8 protein target yang ditargetkan oleh dua senyawa, dan 11 protein target yang ditargetkan oleh satu senyawa. Meskipun memiliki banyak keterbatasan dalam memperoleh performa yang baik, metode profil terdekat dapat digunakan untuk memprediksi target dari senyawa jamu yang belum diketahui targetnya. Namun, pengujian di laboratorium basah masih diperlukan untuk memvalidasi hasil prediksi dengan metode ini.



Gambar 5. *Network* prediksi target dari senyawa formula jamu. Node hijau dan merah merupakan senyawa dari formula jamu. Node hijau merupakan senyawa dari tanaman pare (*Momordica charantia*) dan node merah merupakan senyawa dari tanaman jahe (*Zingiber officinale*). Sementara node biru merupakan protein target dari senyawa.

#### 4. KESIMPULAN

Metode profil terdekat dapat digunakan untuk memprediksi target dari senyawa jamu yang belum diketahui targetnya. Hasil prediksi target dari senyawa jamu, sangat bergantung pada ketersediaan data karakteristik senyawa pada tanaman jamu. Meskipun metode profil terdekat memiliki performa AUC dan akurasi yang baik, namun rendahnya sensitifitas/recall menjadi keterbatasan dalam penelitian ini. Pengujian metode profil terdekat di laboratorium basah masih diperlukan untuk memvalidasi hasil prediksi dengan metode ini.

#### DAFTAR PUSTAKA

- [1] Afendi, F. M., Heryanto, R., Darusman, L. K., Syahrir, N. H. A., Bakri, R., & Qomariasih, N., 2016. Jamu informatics: A new perspective in jamu research. *CICSJ Bulletin*, Vol. 34, No. 2, 47.
- [2] Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, L. K., Saito, K., & Kanaya, S., 2012. KNApSACk family databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology*, Vol. 53, No. 2, 1–12. <https://doi.org/10.1093/pcp/pcr165>
- [3] Bajusz, D., Rácz, A., & Héberger, K., 2015. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics*, Vol. 7,

- No. 1, 1–13. <https://doi.org/10.1186/s13321-015-0069-3>
- [4] Bleakley, K., & Yamanishi, Y., 2009. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics*, Vol. 25, No. 18, 2397–2403. <https://doi.org/10.1093/bioinformatics/btp433>
- [5] Elfahmi, Woerdenbag, H. J., & Kayser, O., 2014. Jamu: Indonesian traditional herbal medicine towards rational phytopharmacological use. *Journal of Herbal Medicine*, Vol. 4, No. 2, 51–73. <https://doi.org/10.1016/j.hermed.2014.01.002>
- [6] Florkowski, C. M., 2008. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *The Clinical Biochemist. Reviews*, 29 Suppl 1(Suppl 1), S83–S87. <https://www.ncbi.nlm.nih.gov/pubmed/18852864>
- [7] Fukunishi, Y., & Nakamura, H., 2011. Prediction of ligand-binding sites of proteins by molecular docking calculation for a random ligand library. *Protein Science*, Vol. 20, No.1, 95–106. <https://doi.org/10.1002/pro.540>
- [8] Liu, Y., Hong, Y., Lin, C. Y., & Hung, C. L., 2015. Accelerating Smith-Waterman Alignment for Protein Database Search Using Frequency Distance Filtration Scheme Based on CPU-GPU Collaborative System. *International Journal of Genomics*, 2015. <https://doi.org/10.1155/2015/761063>
- [9] Tung, C.-W., 2015. Public Databases of Plant Natural Products for Computational Drug Discovery. *Current Computer Aided-Drug Design*, Vol. 10, No. 3, 191–196. <https://doi.org/10.2174/1573409910666140414145934>
- [10] Wang, Y., Bryant, S. H., Cheng, T., Wang, J., Gindulyte, A., Shoemaker, B. A., Thiessen, P. A., He, S., & Zhang, J., 2017. PubChem BioAssay: 2017 update. *Nucleic Acids Research*, Vol. 45, No. D1, D955–D963. <https://doi.org/10.1093/nar/gkw1118>
- [11] Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., MacIejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., ... Wilson, M., 2018. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, Vol. 46, No. D1, D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
- [12] Wu, Z., Li, W., Liu, G., & Tang, Y., 2018. Network-based methods for prediction of drug-target interactions. *Frontiers in Pharmacology*, Vol.9, 1–14. <https://doi.org/10.3389/fphar.2018.01134>
- [13] Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M., 2008. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, Vol. 24, No.23, 232–240. <https://doi.org/10.1093/bioinformatics/btn162>