

# PENINGKATAN AKURASI KLASIFIKASI INTERAKSI FARMAKODINAMIK OBAT BERBASIS SELEKSI PASANGAN OBAT TAKBERINTERAKSI\*

Hilma M. Winata<sup>1</sup>, Farit M. Afendi<sup>2‡</sup>, and Anwar Fitrianto<sup>3</sup>

<sup>1</sup>Department of Statistics, IPB University, Indonesia, hilmamutiarawinata@gmail.com

<sup>2</sup>Department of Statistics, IPB University, Indonesia, fmafendi@apps.ipb.ac.id

<sup>3</sup>Department of Statistics, IPB University, Indonesia, anwarstat@gmail.com

‡corresponding author

**Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)**

**Vol 3 No 3 (2019), 247-259**

Copyright © 2019 Hilma M. Winata, Farit M. Afendi, and Anwar Fitrianto. 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**

Identifying the pharmacodynamics drug-drug interaction (PD DDI) is needed since it can cause side effects to patients. There are two measurements of drug interaction performance, namely the golden standard positive (GSP) which is the drug pairs that interact pharmacodynamics and golden standard negative (GSN), which is a drug pairs that do not interact. The selection of GSN in the previous which studies were only selected randomly from a list of drug pairs that do not interact. The random selection is feared to contain drug pairs that actually interact but have not been recorded. Therefore, in this study the determination of GSN was carried out by, first, grouping drug pairs included in the GSP using the DP-Clus algorithm with certain values of density and cluster properties. Then the drugs in different group would be paired and only the drug pairs in the GSN list are selected. It was found that our new proposed classification method increases the AUC value compared to the results obtained by random selection of GSN.

**Keywords:** DP-Clus, golden standard negative, pharmacodynamics drug-drug interaction.

## **1. Pendahuluan**

Sakit akibat gangguan metabolisme dalam tubuh dapat diakibatkan karena kerja protein dalam tubuh yang terlalu aktif atau terlalu pasif (Evelyn, 2017). Hal ini dapat distabilkan dengan cara mengkonsumsi obat. Jika diharuskan mengkonsumsi lebih

---

\* Received Jun 2019; Accepted Aug 2019; Published online on Oct 2019

dari satu obat maka berpotensi terjadinya interaksi obat yang mungkin memiliki efek samping berbahaya bagi tubuh manusia sehingga harus diketahui bagaimana kondisi interaksinya (Rahmawati *et al.*, 2006).

Mekanisme interaksi obat dapat dibagi menjadi interaksi farmakokinetik dan interaksi farmakodinamik (Fradgley, 2003). Penelitian sebelumnya Huang *et al.* (2013) memprediksi terkait interaksi obat-obat melalui jaringan interaksi protein-protein dimana penelitiannya berfokus pada obat yang berinteraksi farmakodinamik. Interaksi farmakodinamik cukup berbahaya karena efek samping langsung terasa oleh tubuh manusia. Sehingga hal ini perlu untuk diteliti lebih jauh dan efek farmakodinamik dapat diramalkan jika diketahui mekanisme dari kerja obat yang bersangkutan. Penelitian Huang *et al.* (2013) juga mengkategorikan ukuran kinerja pasangan obat menjadi *golden standard positive* (GSP) yang merupakan pasangan obat yang berinteraksi PD dan *golden standard negative* (GSN) yang merupakan pasangan obat tidak berinteraksi. GSP didapat dari suatu data base yang menyediakan informasi terkait pasangan obat mana saja yang berinteraksi, sedangkan GSN hanya dipilih secara acak dari pasangan obat yang bukan termasuk berinteraksi. Pemilihan GSN secara acak dikhawatirkan mengandung pasangan obat yang sebenarnya berinteraksi tapi belum terdata dalam data base yang disediakan. Hampir dibanyak penelitian tidak melakukan seleksi pemilihan GSN dan semua menggunakan GSN yang dipilih secara acak. Sehingga dibutuhkan suatu cara untuk meminimalisir kesalahan dalam menentukan pasangan obat yang termasuk GSN.

Pemilihan GSN disiasati dengan melihat pola interaksi dari obat. Satu obat berinteraksi cenderung dengan suatu obat tertentu. Oleh karenanya pemilihan GSN dapat dilakukan dengan memilih pasangan obat dengan pola interaksi yang berbeda.

Altaf-UI-Amin *et al.* (2007) membuat suatu algoritma pengelompokan grafik berbasis kepadatan pinggiran untuk mendekteksi kompleks protein dalam jaringan interaksi yang bernama DP-Clus. Algoritma ini cenderung mengisolasi daerah yang terhubung secara padat dari sebuah grafik sebagai suatu kelompok. Algoritma ini dapat diterapkan diberbagai jaringan dan di banyak aplikasi lainnya, salah satunya diterapkan dalam obat-obat yang berinteraksi untuk menemukan kelompok yang kohesif.

Pengelompokan menggunakan algoritma DP-Clus dirasa dapat mengatasi kekhawatiran terkait pemilihan GSN secara acak, dengan cara mengelompokkan obat berdasarkan pasangan obat yang berinteraksi PD sehingga menghasilkan karakteristik tersendiri dari setiap kelompok obat. Hal tersebut sesuai dengan tujuan dari penelitian ini yaitu menentukan pasangan obat yang termasuk kedalam GSN dengan mengelompokkan terlebih dahulu pasangan obat yang berinteraksi PD menggunakan DP-CLus. Selanjutnya dianalisis status interaksi pasangan obat dengan metode klasifikasi random forest untuk mengevaluasi akurasi ketepatan dalam prediksinya dan dibandingkan dengan hasil analisis yang GSNnya diambil secara acak.

Komposisi pasangan obat yang termasuk kedalam GSN akan dibangkitkan dari pasangan obat yang berbeda kelompok. Seleksi GSN dengan terlebih dahulu mengelompokkan obat yang termasuk kedalam GSP yang diharapkan dapat

meminimalisir adanya obat yang sebenarnya berinteraksi tetapi belum terdata karena pasangan obat yang dibangun ini dipilih dari 2 kelompok obat yang berbeda. Hasil dari pembentukan GSN kemudian dilihat status interaksi pasangan obatnya dengan klasifikasi random forest, diperoleh akurasi ketepatan dalam prediksinya berdasarkan nilai AUC. AUC adalah cara yang cukup efektif untuk merangkum keakuratan diagnostik keseluruhan tes (Gorunescu, 2011). Nilai AUC disinilah yang dibandingkan dengan hasil analisis yang pemilihan GSNnya diambil secara acak.

## 2. Metodologi

### 2.1 Data

Data yang digunakan dalam penelitian ini adalah data pasangan obat dimana data obat tersebut diambil berdasarkan ID DrugBank (<https://www.drugbank.ca/>) yang disetujui oleh *Food and Drug Administration* (FDA) dan protein targetnya yang ada dalam manusia. Dari DrugBank diambil data pasangan obat yang saling berinteraksi dan dilihat deskripsi interaksinya. Dengan menggunakan metode penambahan teks semi-otomatis diperoleh pasangan obat mana saja yang berinteraksi PD (GSP).

Pada data set digunakan 10546 pasangan obat yang berinteraksi PD dan kurang lebih empat kali lipat pasangan obat yang termasuk kedalam GSN. Data yang digunakan pun terdiri dari satu peubah respon yaitu kondisi interaksinya, GSP atau GSN dan 3 peubah penjelas yaitu *Target Protein Connectedness* (TPC) yang datanya bersumber dari HPRD, *Side Effect Similarity* (SES) yang datanya bersumber dari Meddra dan *Chemical Similarity* (CS) yang datanya bersumber dari PubChem dan Chemdes. Sumber data terkait peubah penjelas semuanya akan di hubungan dengan data pada DrugBank.

### 2.2 Prosedur Analisis Data

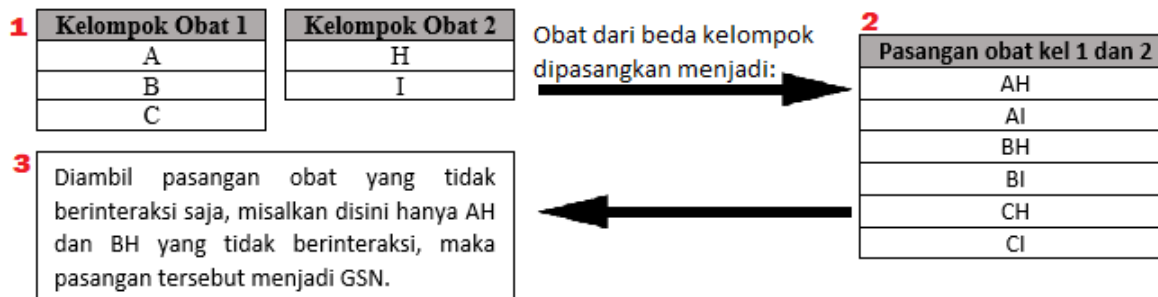
Prosedur analisis data yang akan dilakukan pada penelitian ini adalah sebagai berikut:

1. Menyeleksi data yang digunakan dengan mengumpulkan terlebih dahulu seluruh data yang dibutuhkan dari berbagai sumber data yaitu data pasangan obat dengan keterangan interaksi dari pasangan obat, nilai TPC, nilai SES dan nilai CS.
  - a) Mengumpulkan data terkait obat-obat yang digunakan, target protein dari setiap obat dan pasangan obat yang berinteraksi dengan melakukan scraping data terlebih dahulu.
  - b) Menghubungkan data dari DrugBank dengan data dari HPRD terkait dengan interaksi protein-protein yang selanjutnya dilakukan perhitungan TPC. TPC merupakan suatu nilai yang menggambarkan keterhubungan protein target antar obat. Perhitungan TPC dilakukan dengan melihat protein target dari pasangan obat yang akan dihitung, kemudian setiap protein target diambil subjaringan dari jaringan interaksi protein-protein yang ada pada seluruh tubuh. Subjaringan itu adalah jaringan protein target dengan protein lainnya yang bertetangga tepat satu dengan protein target tersebut. Setiap subjaringan dari pasangan obat tersebut dihitung nilai banyaknya protein tetangga, simpangan baku dan rata-rata dari bobot yang diperoleh berdasarkan koefisien

korelasi pearson (PCC) dari pengkodean profil ekspresi gen di 79 jaringan manusia. Kemudian dihitung rata-rata PCC dari semua titik dalam jaringan. Jika terdapat pasangan obat yang memiliki target protein yang sama maka ditambah 1 diantara kedua sistem tersebut. Hal ini berulang untuk setiap pasangan obat yang dipilih untuk dihitung nilai TPCnya menggunakan rumus dibawah ini:

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

- c) Mengumpulkan data efek samping obat dari medra kemudian dihubungkan dengan ID obat dari DrugBank menggunakan data obat yang terlebih dahulu disambungkan melalui data obat di Matador dan STITCH, sehingga nilai SES dapat dihitung. Nilai ini merupakan penggambaran terkait ukuran kesamaan efek samping dari pasangan obat (Campillos *et al.*, 2008).
  - d) Mengumpulkan data fingerprint dari Chemdes yang selanjutnya diubah kedalam bentuk SMILES (bentuk molekulnya) menggunakan PubChem (Dong *et al.*, 2015) dan dihitung nilai dari fungsi kemiripan tanimoto (Todeschini *et al.*, 2012). Perhitungan ini menghasilkan nilai CS yaitu ukuran kesamaan struktur molekul dari setiap pasangan obat tersebut.
2. Data pasangan obat yang termasuk kedalam GSP dikelompokkan menggunakan algoritma DP-Clus yang dikembangkan oleh Altaf-UI-Amin *et al.* (2007). Sistem kerja dalam algoritma ini adalah mendeteksi daerah yang terhubung secara padat dari grafik yang dipisahkan oleh daerah yang jarang berdasarkan nilai kepadatan ( $0 < d_{in} \leq 1$ ) dan nilai *cluster property* ( $0 < cp_{in} \leq 1$ ) tertentu. Pada penelitian ini dicobakan dengan 2 kondisi.
    - a) Kondisi 1, ketika nilai kepadatannya dibuat tetap yaitu 0.6 dan nilai *cluster property*nya dibuat berubah-ubah dimulai dari 0.1 sampai 0.9 dengan peningkatan urutan sebesar 0.1.
    - b) Kondisi 2, ketika nilai *cluster property*nya dibuat tetap yaitu 0.5 dan nilai kepadatannya dibuat berubah-ubah dimulai dari 0.1 sampai 0.9 dengan peningkatan urutan sebesar 0.1.
 Pemilihan nilai kepadatan ( $d = 0.6$ ) dan *cluster property* ( $cp = 0.5$ ) yang dibuat tetap dipilih berdasarkan rujukan penelitian Altaf-UI-Amin *et al.* (2007).
  3. Eksplorasi data untuk setiap kelompok obat yang terbentuk menggunakan DP-Clus dengan pertama-tama memasang setiap obat pada kelompok yang sama dan mengambil pasangan obat yang termasuk kedalam GSP saja untuk dilihat kondisi dari nilai TPC, SES dan CS berdasarkan pasangan obat yang terbentuk.
  4. Memasang obat dari beda kelompok yang tidak berinteraksi untuk dijadikan golden standard negatives (GSN) seperti pada Gambar 1.
  5. Lakukan metode klasifikasi *random forest* untuk melihat nilai AUC yang dimiliki dengan GSN yang baru terbentuk dengan melakukan *cross validation* dimana banyak folds adalah 5. Bandingkan nilai AUC dataset dengan GSN yang baru terbentuk dengan GSN yang dipilih secara acak seperti pada penelitian sebelumnya.



Gambar 1: Algoritma penentuan GSN.

Tabel 1: Informasi pangkalan data yang dirujuk untuk pembentukan dataset.

Jenis Data	Pangkalan Data	Informasi yang diperoleh dari pangkalan data	Hasil penelusuran data
Obat	DrugBank <a href="https://www.drugbank.ca/">https://www.drugbank.ca/</a>	Informasi rinci terkait ID obat DrugBank, obat-obat yang saling berinteraksi, obat dan protein targetnya.	Terdapat 5598 obat yang disetujui FDA, dengan terdiri dari 106939 pasangan obat yang berinteraksi dimana 69009 diantaranya berinteraksi PD
Protein	Human Protein Reference Database (HPRD) <a href="http://www.hprd.org/">http://www.hprd.org/</a>	Informasi terkait protein-protein yang saling berinteraksi dan protein yang diketahui <i>gene expression profil</i> nya	Terdapat 30005 protein-protein berinteraksi yang diketahui bobotnya dan 5445 protein yang diketahui <i>gene expression profil</i> nya
Integrasi pangkalan data	STITCH <a href="http://stitch.embl.de/">http://stitch.embl.de/</a>	Informasi terkait ID senyawa STITCH.	Konversi ID protein target pada STITCH berupa "chemical ID" menjadi "protein ID".
Integrasi pangkalan data	STITCH <a href="http://stitch.embl.de/">http://stitch.embl.de/</a>	Informasi terkait ID senyawa STITCH.	Konversi ID protein target pada STITCH berupa "chemical ID" menjadi "protein ID".
Integrasi pangkalan data	Matador <a href="http://matador.embl.de/">http://matador.embl.de/</a>	Informasi terkait protein ID.	Konversi ID protein target pada Matador berupa "protein ID" menjadi "uniprot ID"
Jenis Data	Pangkalan Data	Informasi yang diperoleh dari pangkalan data	Hasil penelusuran data
Efek samping obat	Meddra <a href="https://www.meddra.org/">https://www.meddra.org/</a>	Informasi rinci terkait efek samping dari setiap obat.	Terdapat 1030 obat yang diketahui efek sampingnya dan diketahui ID DrugBanknya
Bahan Aktif	PubChem <a href="https://pubchem.ncbi.nlm.nih.gov">https://pubchem.ncbi.nlm.nih.gov</a>	Informasi mengenai bahan aktif berupa struktur kimia (canonical SMILES) dan PubChem ID (CID).	Terdapat 5254 obat yang diketahui struktur kimianya
Klekota-Roth Fingerprint	ChemDes <a href="http://www.scbdd.com/chemdes/">http://www.scbdd.com/chemdes/</a>	Daftar kriteria <i>fingerprint</i> .	Terdapat 704 obat yang diketahui <i>fingerprint</i> nya

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Penyeleksian Data

Pengumpulan data dilakukan dengan cara menelusuri beberapa pangkalan data yang diperoleh dari berbagai publikasi ilmiah yang terkait. Pangkalan data tersebut dihimpun untuk dijadikan satu data set baru. Informasi terkait pangkalan data dideskripsikan seperti pada Tabel 1.

Pengumpulan data yang bersumber dari berbagai pangkalan data menghasilkan irisan yang kemudian digunakan sebagai data set dalam penelitian ini. Data set tersebut terdiri dari 654 obat yang sudah disetujui FDA.

Data set yang terbentuk terdiri dari 3 peubah penjelas yang dibangun berdasarkan hasil perhitungan-perhitungan tertentu. Berdasarkan hasil perhitungan tersebut diperoleh nilai TPC, SES dan CS pada setiap pasangan obat yang terbentuk dengan terdiri dari 10546 pasangan obat yang termasuk GSP dan 202985 pasangan obat yang diindikasikan sebagai GSN. Kemudian, analisis lanjut terkait klasifikasi yang akan dilakukan hanya menggunakan data sebanyak kurang lebih 4 kali lipat data yang akan dijadikan pasangan obat yang termasuk kedalam GSN. Ringkasan terkait nilai dari 3 peubah penjelas yang ada pada dalam data set disajikan seperti Tabel 2.

Tabel 2: Ringkasan nilai TPC, SES dan CS.

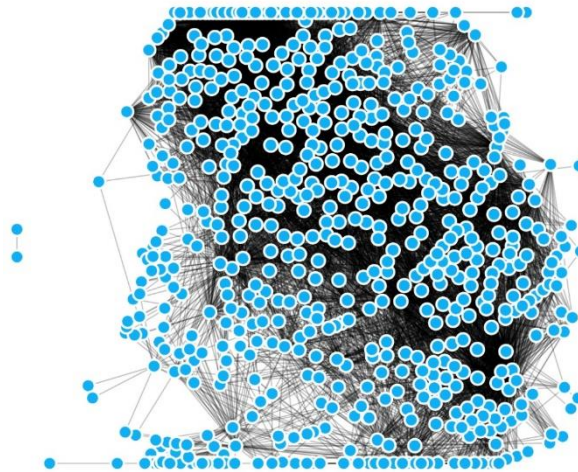
	TPC	SES	CS
Min.	-24.7030	0.0000	0.0000
1st Qu.	0.1277	0.0000	0.1000
Median	2.5620	0.0000	0.1522
Mean	2.1356	5.8840	0.1651
3rd Qu.	4.5766	0.0000	0.2162
Max.	400.9049	830.9060	1.0000

Tabel 2 memberikan informasi terkait kondisi peubah penjelas yang digunakan dalam analisis. Nilai TPC dan SES memiliki range data yang jauh. Sedangkan nilai CS merupakan nilai peluang sehingga berada pada nilai 0 sampai 1. Nilai TPC yang semakin tinggi artinya pasangan obat tersebut semakin dekat keterhubungannya dilihat berdasarkan protein targetnya. Nilai SES 0 itu berarti pasangan obat tersebut tidak memiliki efek samping yang sama dan semakin tinggi nilai SES maka semakin banyak kesamaan efek samping antar pasangan obat tersebut.

Jika dilihat keterhubungan antar peubah seperti pada Tabel 3, antar peubah penjelas tersebut memiliki hubungan yang lemah. Selain itu, setiap peubah penjelas memiliki keterhubungan dengan peubah respon walau hubungannya juga lemah. Hal ini diharapkan peubah penjelas dapat saling melengkapi dalam memberikan informasi terhadap peubah respon.

Tabel 3: Korelasi antar peubah.

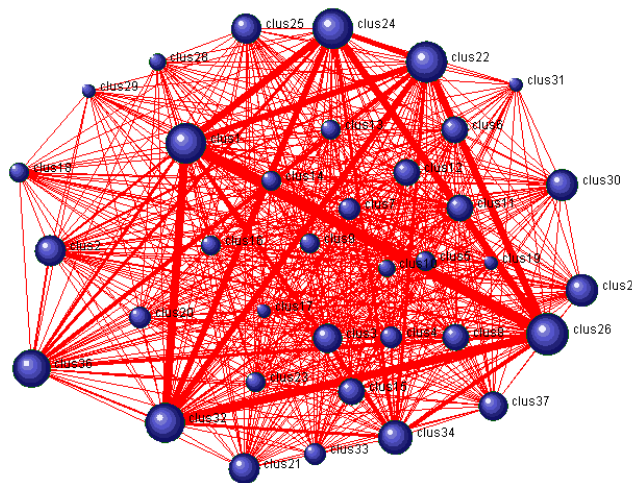
	TPC	SES	CS
TPC	1.000	-0.007	-0.029
SES	-0.007	1.000	0.022
CS	-0.029	0.022	1.000
kondisi interaksi (Y)	0.109	0.033	0.011



Gambar 2: Jejaring pasangan obat yang termasuk kedalam GSP.

### 3.2 Pembentukan dan eksplorasi GSN

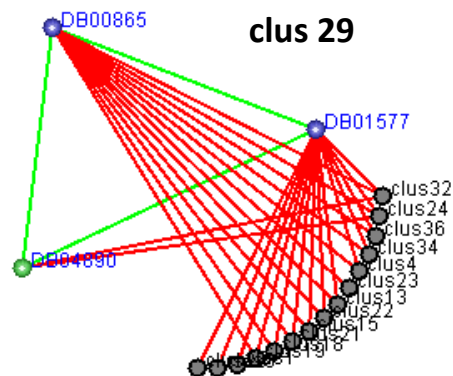
Data yang terbentuk sebagai GSP selanjutnya digunakan dalam algoritma DP-Clus. Gambar 2 merupakan jejaring dari pasangan obat yang termasuk kedalam GSP. Jejaring yang besar ini akan dipecah dan dibentuk kelompok-kelompok berdasarkan nilai *cluster property* dan kepadatan yang dipilih.



Gambar 3: Jejaring kelompok obat dengan  $Cp = 0.9$  dan  $d = 0.6$ .

Jejaring yang besar yang memuat seluruh pasangan obat yang termasuk GSP akan dibentuk menjadi kelompok-kelompok tertentu. Gambar 3 merupakan salah satu contoh jejaring yang terbentuk dengan nilai *cluster property* ( $cp = 0.9$ ) dan kepadatan ( $d = 0.6$ ). Berdasarkan nilai *cluster property* dan kepadatan tersebut terbentuklah sebanyak 37 kelompok yang jika dilihat lebih dalam lagi bentuk jejaring dari setiap kelompoknya tergambarakan seperti pada Gambar 4.

Pengelompokan menggunakan DP-Clus dengan nilai kepadatan dan cluster property tertentu menghasilkan jumlah kelompok yang berbeda-beda, sehingga banyak GSN yang terbentuk juga tidak sama pada setiap kondisinya. Hasil yang diperoleh dari pengelompokan tersebut dapat dilihat seperti pada Tabel 4.



Gambar 4: Jejaring obat pada kelompok 29 dimana  $Cp = 0.9$  dan  $d = 0.6$ .

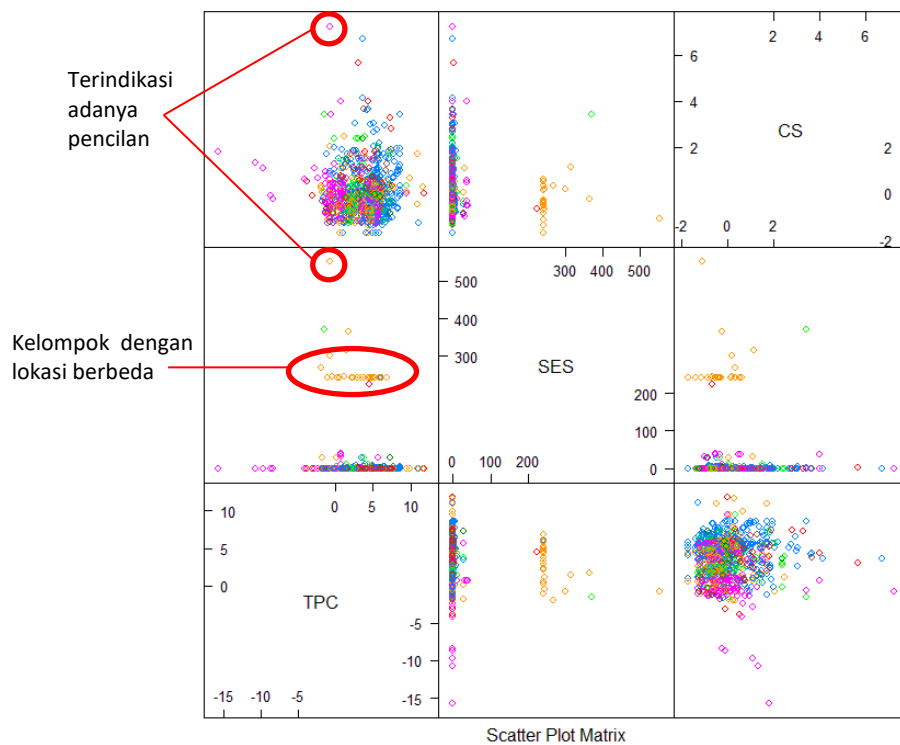
Kondisi 1 dimana nilai cluster property dibuat tetap dan nilai kepadatan lebih dari 0.6 membentuk GSN yang cenderung seimbang dengan banyak pasangan GSP yang dimiliki. Sedangkan kondisi lainnya membentuk GSN yang jumlah kurang lebih 5 kali lebih banyak dari GSP yang dimiliki.

Pengelompokan berdasarkan algoritma DP-Clus salah satunya digunakan untuk melihat kondisi dari pasangan obat yang termasuk kedalam GSP yang dibangun dari obat pada kelompok yang sama. Hal ini dilakukan untuk dimanfaatkan dalam pembentukan pasangan obat yang termasuk ke dalam GSN. Kondisi dari pasangan obat yang termasuk kedalam GSP yang dibangun dari obat pada kelompok yang sama dilihat berdasarkan nilai TPC, SES dan CS yang terbentuk Gambar 5.



Tabel 4: Banyak GSN dan kelompok yang terbentuk dengan kepadatan dan cluster property tertentu.

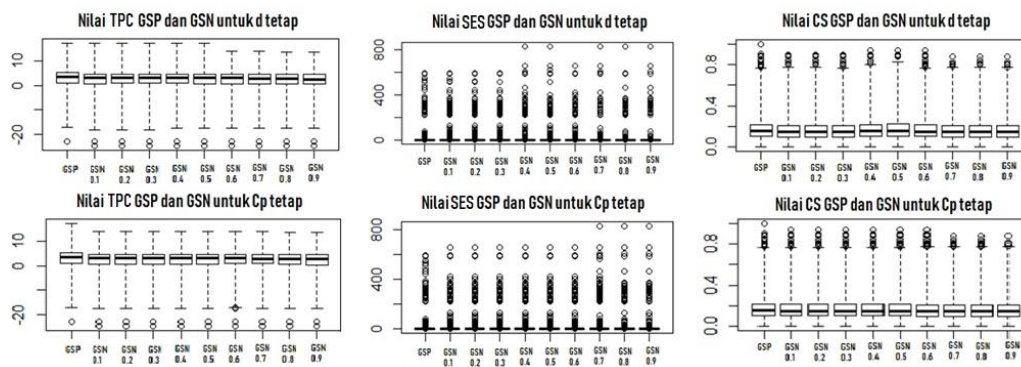
GSN yang terbentuk	Kelompok	cp	d
51229	17	0.5	0.1
49074	17	0.5	0.2
48883	19	0.5	0.3
45798	21	0.5	0.4
46126	24	0.5	0.5
48593	27	0.5	0.6
19595	31	0.5	0.7
16560	34	0.5	0.8
13157	37	0.5	0.9
49170	27	0.1	0.6
49170	27	0.2	0.6
49170	27	0.3	0.6
48890	27	0.4	0.6
48593	27	0.5	0.6
73852	27	0.6	0.6
64309	28	0.7	0.6
56826	34	0.8	0.6
55203	37	0.9	0.6



Gambar 5: Plot matriks untuk kelompok yang terbentuk pada Cp=0.9 dan d=0.6.

Berdasarkan Gambar 5 dapat dilihat kondisi pasangan obat berinteraksi PD yang terbentuk berdasarkan nilai TPC, SES dan CS. Plot ini menggambarkan hubungan untuk setiap 2 peubah yang digunakan dalam analisis. Warna pada gambar merupakan pembeda antar 37 kelompok yang terbentuk pada  $C_p=0.9$  dan  $d=0.6$ . Terlihat pada plot untuk peubah TPC dan CS data tidak membentuk suatu hubungan tertentu dan setiap kelompok tidak tampak perbedaannya. Sedangkan pada plot untuk peubah TPC dengan SES dan SES dengan CS masing-masingnya mengandung satu kelompok yang cukup jelas berada pada beda lokasi. Kelompok yang berbeda lokasi itu memiliki nilai SES lebih tinggi dibandingkan dengan kelompok lainnya secara konstan. Gambar 3 juga menginformasikan bahwa terdapat beberapa data yang diindikasikan sebagai pencilan seperti titik-titik yang sangat jauh dari kelompok.

Obat dari beda kelompok dipasangkan dan dipilih pasangan obat yang termasuk kedalam GSN untuk masing-masing kondisi dilihat kondisi berdasarkan peubah penjelas yang digunakan dalam analisis seperti pada Gambar 6.



Gambar 6: Boxplot pada setiap peubah pada dua kondisi.

Pada Gambar 6 untuk boxplot “Nilai TPC GSP dan GSN untuk  $d$  tetap” maksudnya adalah kondisi nilai TPC pada kelompok GSP dan kelompok GSN yang dibentuk dengan nilai kepadatan ( $d$ ) tetap dan nilai *cluster property* yang berubah-ubah dari 0.1-0.9 dengan peningkatan urutan sebesar 0.1. Terlihat bahwa dari ke enam gambar yang terbentuk tidak begitu terlihat perbedaan dari setiap boxplot yang terbentuk. Ini berarti bahwa kondisi setiap peubah pada pasangan obat yang termasuk GSP maupun pasangan obat yang termasuk kedalam GSN yang dibentuk berdasarkan nilai kepadatan dan cluster property tertentu sulit untuk dibedakan.

Berdasarkan matriks plot pada Gambar 5 hanya terdapat 1 kelompok yang terlihat jelas perbedaannya dengan kelompok lain yaitu berdasarkan peubah TPC dengan SES dan SES dengan CS. Jika dilihat dari pembentukan GSN pada nilai kepadatan dan nilai cluster property tertentu dan dibandingkan juga dengan pasangan obat yang termasuk kedalam GSP tidak terlihat jelas bagaimana perbedaan berdasar nilai peubah TPC, SES dan CS.

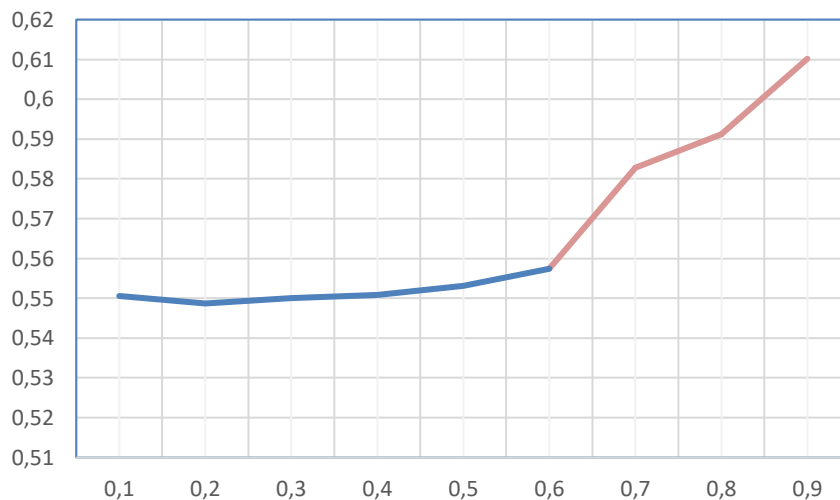
### 3.3 Peforma Hasil Klasifikasi

Hasil dari pembentukan pasangan obat yang termasuk GSN berdasarkan algoritma DP-Clus diklasifikasikan menggunakan metode *random forest* yang kemudian

menghasilkan nilai AUC yang berbeda-beda pada setiap kondisinya seperti pada Tabel 5.

Tabel 5: Nilai AUC dengan cluster property dan kepadatan tertentu

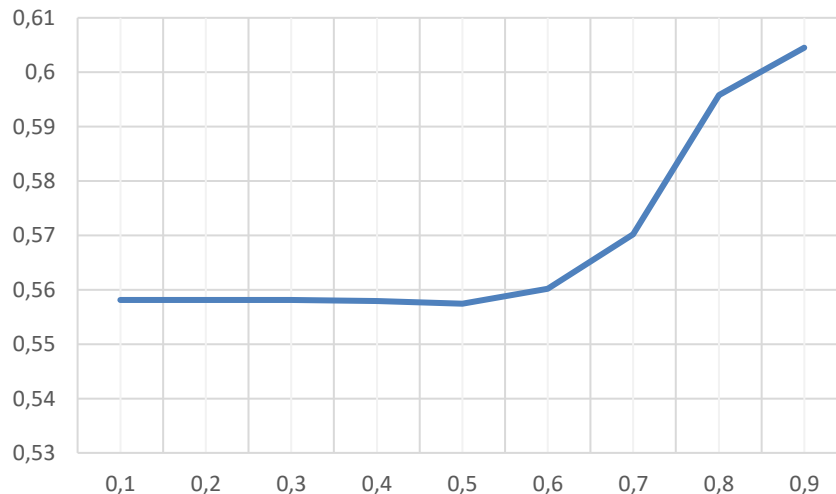
cp	d	AUC
0.5	0.1	0.55056
0.5	0.2	0.54868
0.5	0.3	0.54998
0.5	0.4	0.55082
0.5	0.5	0.55315
0.5	0.6	0.55743
0.5	0.7	0.58278
0.5	0.8	0.59119
0.5	0.9	0.61016
0.1	0.6	0.55817
0.2	0.6	0.55817
0.3	0.6	0.55817
0.4	0.6	0.55795
0.5	0.6	0.55743
0.6	0.6	0.56020
0.7	0.6	0.57023
0.8	0.6	0.59583
0.9	0.6	0.60451



Gambar 7: Plot AUC ketika cluster property tetap ( $Cp = 0.5$ ).

Gambar 7 merupakan penggambaran untuk nilai AUC pada kondisi nilai *cluster property* yang dibuat tetap yaitu  $cp = 0.5$ . Terlihat bahwa nilai AUC mengalami peningkatan terutama ketika kepadatannya 0.7-0.9. Peningkatan nilai AUC pada kondisi tersebut dikarenakan kondisi dataset yang cenderung seimbang antara GSP

dan GSN. Karena pada kondisi nilai kepadatan 0.7-0.9 GSN yang terbentuk masing-masing hanya 19595, 16560 dan 13157 pasang. Yang jumlahnya tidak terlalu jauh dengan banyak pasangan yang termasuk kedalam GSP yaitu 10546. Karena kondisinya yang hampir seimbang maka klasifikasi *random forest* akan menghasilkan perhitungan ketepatan memprediksi negatif yang juga benar negatif dan memprediksi positif yang juga benar positif hampir seimbang sehingga menghasilkan nilai AUC yang cukup tinggi dibandingkan dengan kondisi lainnya.



Gambar 8: Plot AUC ketika kepadatan tetap ( $d=0.6$ ).

Gambar 8 merupakan penggambaran untuk nilai AUC pada kondisi nilai kepadatan yang dibuat tetap yaitu  $d = 0.6$ . Terlihat bahwa nilai AUC mengalami peningkatan terutama ketika nilai *cluster property*-nya 0.7-0.9. Berdasarkan penggambaran tersebut semakin besar nilai *cluster property*-nya dengan nilai kepadatan yang dibuat tetap maka semakin tinggi nilai AUC.

Nilai AUC yang dihasilkan dari klasifikasi dengan pengambilan GSN secara acak tanpa mengelompokkannya terlebih dahulu adalah sebesar 0.5608. Jika dibandingkan, nilai dari kedua kondisi baik yang kondisi 1 dimana nilai *cluster property* tetap menghasilkan nilai AUC yang lebih tinggi dibandingkan dengan nilai AUC hasil klasifikasi dengan GSN secara acak ketika nilai kepadatan lebih dari 0.7 dan kondisi 2 dimana nilai kepadatan tetap menghasilkan nilai AUC yang lebih tinggi dibandingkan dengan nilai AUC hasil klasifikasi dengan GSN secara acak ketika nilai *cluster property*-nya lebih dari 0.7.

#### 4. Simpulan

Penerapan algoritma DP-Clus pada pasangan obat yang berinteraksi PD (GSP) dibagi menjadi dua kondisi. Pada kondisi 1 dimana nilai kepadatan dibuat tetap yaitu  $d = 0.6$  dan nilai *cluster property*-nya berubah-ubah dimulai dari 0.1 sampai 0.9 dengan peningkatan urutan sebesar 0.1 menghasilkan peningkatan nilai AUC. Jika dibandingkan dengan nilai AUC dengan GSN yang diambil secara acak maka diperoleh nilai AUC yang lebih tinggi ketika nilai *cluster property*-nya lebih dari 0.7. Pada kondisi 2 dimana nilai *cluster property* dibuat tetap yaitu  $cp = 0.5$  dan nilai

kepadatannya berubah-ubah dimulai dari 0.1 sampai 0.9 dengan peningkatan urutan sebesar 0.1 menghasilkan peningkatan nilai AUC. Jika dibandingkan dengan nilai AUC dengan GSN yang diambil secara acak maka diperoleh nilai AUC yang lebih tinggi ketika nilai *cluster property*nya lebih dari 0.7.

Berdasarkan dua kondisi tersebut dan menjawab tujuan dari penelitian ini. Pembentukan pasangan obat yang termasuk GSN dengan terlebih dahulu menerapkan algoritma DP-Clus pada pasangan obat yang termasuk kedalam GSP, agar memperoleh nilai akurasi ketepatan prediksi yang lebih baik dapat menerapkan dua kondisi yaitu ketika nilai kepadatan yang dibuat tetap maka gunakan nilai *cluster property* yang tinggi begitu pula jika nilai *cluster property*nya dibuat tetap maka gunakan nilai kepadatan yang tinggi.

### Daftar Pustaka

- Altaf-UI-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.
- 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): 60. <https://doi.org/10.1186/s13321-015-0109-z>.
- Evelyn, T. (2017, February 13). *Berbagai Penyakit Akibat Gangguan Metabolisme*. Retrieved from <https://health.uzone.id/berbagai-penyakit-akibat-gangguan-metabolisme>.
- Fradgley, S. (2003). *Interaksi Obat dalam Farmasi Klinis (Clinical Pharmacy) Menuju Pengobatan Rasional dan Penghargaan Pilihan Pasien (Aslam M, Tan CK, Prayitno A, Ed)*. Jakarta (ID): PT Elex Media Komputindo Kelompok Gramedia.
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). German: Springer Science & Business Media.
- 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.
- Rahmawati, F., Handayani, R., & Gosal, V. (2006). Kajian retrospektif interaksi obat di Rumah Sakit Pendidikan Dr. Sardjito Yogyakarta. *Majalah Farmasi Indonesia*, **17**(4): 177 – 183.
- Todeschini, R., Consonni, V., Xiang, H., Holliday, J., Buscema, M., & Willett, P. (2012). Similarity coefficients for binary chemoinformatics data: Overview and extended comparison using simulated and real data sets. *Journal of Chemical Information and Modeling*, **52**(11): 2884–2901.