

**Penerapan Optimasi Portfolio Untuk Pemilihan Kandidat  
Molekul Dalam Menghambat Ptp1b Pada Penyakit Diabetes  
Melitus Menggunakan Non-dominated Sorting Genetic Algorithm  
(NSGA-II)**

**Tugas Akhir  
diajukan untuk memenuhi salah satu syarat  
memperoleh gelar sarjana  
dari Program Studi Sarjana Informatika  
Fakultas Informatika  
Universitas Telkom**

**1301160792  
Rafanzhani Elfarizy**



**Program Studi Sarjana Informatika  
Fakultas Informatika  
Universitas Telkom  
Bandung  
2020**

## LEMBAR PENGESAHAN

**Penerapan Optimasi Portfolio Untuk Pemilihan Kandidat Molekul Dalam Menghambat Ptp1b Pada Penyakit Diabetes Melitus Menggunakan Non-dominated Sorting Genetic Algorithm (NSGA-ii)**

*Application Of Portfolio Optimization For Selection Of Molecular Candidates In Inhibiting Ptp1b In Diabetes Mellitus Using Non-dominated Sorting Genetic Algorithm (NSGA-ii)*

**NIM : 1301160792**

**Rafanzhani Elfarizy**

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh gelar pada Program Studi Sarjana Informatika  
Fakultas Informatika  
Universitas Telkom

Bandung, 30 Juni 2020

Menyetujui

Pembimbing I,

Pembimbing II,

Dr. Deni Saepudin, S.Si., M.Si.  
99750013

Isman Kurniawan, S.Pd., M.Si., M.Sc., Ph.D.  
15870066

Ketua Program Studi  
Sarjana Informatika

,

Niken Dwi Wahyu Cahyani S.T.,M.Kom,Ph.D  
NIP: 91930541

## LEMBAR PERNYATAAN

Dengan ini saya, Rafanzhani Elfarizy, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul **Penerapan Optimasi Portfolio Untuk Pemilihan Kandidat Molekul Dalam Menghambat Ptp1b Pada Penyakit Diabetes Melitus Menggunakan Non-dominated Sorting Genetic Algorithm (NSGA-II)** beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan. Saya siap menanggung resiko/sanksi yang diberikan jika di kemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam buku TA atau jika ada klaim dari pihak lain terhadap keaslian karya,

Bandung, 30 Juni 2020  
Yang Menyatakan

Rafanzhani Elfarizy

# **Penerapan Optimasi Portfolio Untuk Pemilihan Kandidat Molekul Dalam Menghambat Ptp1b Pada Penyakit Diabetes Melitus Menggunakan Non-dominated Sorting Genetic Algorithm**

**Rafanzhani Elfarizy<sup>1</sup>, Deni Saepudin<sup>2</sup>, Isman Kurniawan<sup>3</sup>**

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>erafanzhani@students.telkomuniversity.ac.id, <sup>2</sup> denisaepudin@telkomuniversity.ac.id,

<sup>3</sup>ismankrn@telkomuniversity.ac.id

---

## **Abstrak**

PTP1B atau Protein Tyrosine Phosphatase 1B merupakan protein yang menjadi salah satu penyebab dari penyakit Diabetes Melitus. Salah satu cara untuk menanganinya adalah dengan menghambat PTP1B. Terdapat banyak kandidat molekul yang berpotensi menghambat pertumbuhan protein ini. Untuk membantu meringankan pemilihan, molekul dalam jumlah yang besar ini dipilih berdasarkan tingkat probabilitas kesuksesan yang tinggi. Masalah pemilihan ini menyerupai masalah pemilihan saham untuk portfolio yang optimal pada keuangan. Permasalahan ini dapat diselesaikan dengan menggunakan NSGA-II berdasarkan prinsip Multi-Objective Optimization. Pada algoritma ini dalam setiap generasi dilakukan evaluasi berdasarkan non-dominated sorting terhadap individu pada populasi untuk mendapatkan individu terbaik yang akan menjadi parent pada generasi selanjutnya. Parent kemudian akan digunakan untuk menghasilkan himpunan turunan (offsprings). Pada akhir generasi akan didapatkan himpunan individu terbaik dengan beberapa molekul yang menjadi anggotanya. Sebanyak 3715 data yang digunakan dalam penelitian ini diambil dari [www.ebi.ac.uk](http://www.ebi.ac.uk). Setelah dilakukan preprocessing terdapat sebanyak 1852 data yang memenuhi klasifikasi untuk dapat digunakan. Pengujian yang dilakukan terhadap dataset sebanyak 30 kali pengujian untuk jumlah molekul dalam 1 portfolio yang berbeda. Jumlah molekul yang diujikan adalah 5, 10 dan 25 molekul dalam 1 portfolio. Hasil yang didapatkan menunjukkan ketiga grafik tersebut memiliki tingkat konvergensi yang kecil.

**Kata kunci :** multi-objective optimization, NSGA-II, expected return, diversity, efficient frontier

---

## **Abstract**

PTP1B or Protein Tyrosine Phosphatase 1B is a protein that is one of the causes of Diabetes Mellitus. One way to handle it is by inhibiting PTP1B. There are many candidate molecules that have the potential to inhibit the growth of this protein. To make selection easier, these large numbers of molecules are chosen based on a high probability of success. This selection problem is similar to the issue of stock selection for an optimal portfolio in finance. This problem can be solved by using NSGA-II based on the principle of Multi-Objective Optimization. In this algorithm, each generation is evaluated based on non-dominated sorting of individuals in the population to get the best individual that will be the parent of the next generation. The parent will then be used to produce offsprings. At the end of the generation, it will get the best set of individuals with several molecules that are members. A total of 3715 data used in this study were taken from [www.ebi.ac.uk](http://www.ebi.ac.uk). After preprocessing there are 1852 data that meet the classification to be used. Tests carried out on the dataset as many as 30 times testing for the number of molecules in a different portfolio. The number of molecules tested was 5, 10 and 25 molecules in 1 portfolio. The results obtained show that the three graphs have a small degree of convergence.

**Keywords:** multi-objective optimization, NSGA-II, expected return, diversity, efficient frontier

---

## **1. Pendahuluan**

### **1.1 Latar Belakang**

Setiap organisme mempunyai peranannya sendiri dalam menyerang sistem tubuh manusia. Salah satu contohnya adalah PTP1B. PTP1B atau Protein tyrosine phosphatase 1B merupakan sebuah jenis protein yang menjadi salah satu penyebab dari penyakit Diabetes Melitus. Langkah yang dapat diambil untuk mengobati pasien dengan penyakit Diabetes Melitus adalah dengan membuat obat yang di dalamnya terkandung molekul yang bisa menghambat pertumbuhan PTP1B [1].

Pada awalnya pemilihan molekul penghambat ini hanya didasari oleh probabilitas kesuksesan dari masing-masing molekul tersebut. Namun, hal ini dinilai tidak cukup, bahkan berisiko. Untuk mengatasi permasalahan tersebut pendekatan alternatif yang dilakukan adalah dengan melihat nilai diversity dari masing – masing molekul. Hal ini dimaksudkan agar untuk setiap kandidat molekul terpilih mengalami kegagalan, molekul lain yang memiliki keberagaman akan memberi peluang lain [2].

Masalah pemilihan molekul tersebut mirip dengan masalah pemilihan saham pada portfolio keuangan. Pada portfolio keuangan, komposisi saham ditentukan berdasarkan nilai expected return tertinggi untuk

suatu tingkat risiko tertentu pada suatu portfolio. Pemilihan portfolio dengan mempertimbangkan 2 fungsi objektif ini dapat diselesaikan dengan metode multi-objective optimization [2].

Pada tugas akhir ini akan diimplementasikan portfolio pada keuangan untuk menyelesaikan masalah pemilihan kandidat molekul yang menghambat pertumbuhan PTP1B. Dalam menyelesaikan masalah ini ada dua fungsi objektif yang harus dimaksimumkan yaitu, expected return yang direpresentasikan sebagai probabilitas kesuksesan molekul dan diversity atau keberagaman molekul. Dengan menggunakan konsep multi-objective akan dihasilkan kemungkinan solusi yang dapat dipilih dengan mempertimbangkan kedua fungsi tersebut. salah satu algoritma untuk yang dapat menyelesaikan permasalahan multi-objective optimization adalah NSGA-II. Algoritma ini akan membentuk kumpulan solusi dengan melihat tingkat dominansi antar solusi. Sehingga tidak ada solusi yang saling mendominasi satu dengan yang lainnya.

## 1.2 Topik dan Batasannya

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, pada Tugas Akhir ini akan dibahas tentang bagaimana cara menentukan komposisi kandidat molekul penghambat PTP1B dengan menggunakan metode optimasi portfolio pada keuangan dengan menerapkan algoritma NSGA-II dalam menyelesaikan permasalahan multi-objective optimization. Adapun batasan masalah yang digunakan pada tugas akhir ini adalah:

1. Dataset yang digunakan diambil dari [www.ebi.ac.uk](http://www.ebi.ac.uk)
2. Jumlah molekul dalam satu portofolio telah ditentukan.
3. Hanya data molekul dengan nilai activity lebih dari 10000 nM yang dapat digunakan.
4. Penggunaan Molecular Fingerprint berdasarkan library rdkit pada python yang merupakan hasil transformasi dari SMILES yang disediakan pada dataset

## 1.3 Tujuan

Penelitian ini bertujuan untuk menghasilkan komposisi kandidat molekul penghambat pertumbuhan PTP1B dengan nilai expected return tertinggi untuk suatu tingkat diversity tertentu dengan menerapkan metode Multi-Objective Optimization pada portfolio keuangan menggunakan algoritma NSGA-II.

## 1.4 Organisasi Tulisan

Urutan penulisan dalam tugas akhir ini adalah sebagai berikut : bagian 2 menunjukkan studi literatur yang terkait dengan penelitian ini. Bagian 3 menjelaskan sistem yang akan dibangun mulai dari preprocessing data untuk menghasilkan data yang dapat diolah sampai dengan mengolah data menggunakan NSGA-II. Bagian 4 menjelaskan Analisa mengenai penelitian yang dilakukan. Bagian 5 menjelaskan kesimpulan yang didapat dan penelitian yang mungkin bisa dikerjakan untuk selanjutnya.

## 2. Studi Terkait

### 2.1 Drug Discovery

PTP1B atau Protein tyrosine phosphatase 1B merupakan jenis protein yang terbukti sebagai negative regulator dalam penjaluran sinyal insulin yang dapat menyebabkan penyakit metabolic seperti obesitas dan diabetes melitus . Protein ini telah dibuktikan berpotensi tinggi untuk menginduksi resistensi insulin dalam tubuh manusia. Hal ini yang menyebabkan PTP1B menjadi target yang tepat untuk menemukan pengobatan yang tepat untuk penyakit diabetes melitus [1].

### 2.2 Portfolio Keuangan

#### 2.2.1 Portfolio

Portfolio sering dikaitkan dengan manajemen investasi. Para investor dalam melakukan investasi pasti selalu mengharapkan keuntungan semaksimal mungkin dengan risiko yang seminimal mungkin atas biaya yang dikeluarkan [3] .

Markowitz membuktikan dalam teorinya bahwa untuk membentuk suatu portfolio yang optimal, diperlukan pembagian modal untuk beberapa sekuritas investasi. [4].

#### 2.2.2 Return dan Risk

Nilai pengembalian dalam portfolio keuangan diartikan sebagai jumlah keuntungan yang didapat dari setiap sekuritas yang dipilih sebagai portfolio. Expected return merupakan rata – rata keuntungan yang didapat untuk suatu portfolio yang dibangun [5].

#### 2.2.3 Bobot Aset

Pada tugas akhir ini, bobot aset direpresentasikan sebagai komposisi pemilihan molekul yang akan dipilih. Pemilihan diputuskan dengan memilih molekul (bobot = 1) atau tidak (bobot = 0) [2].

## 2.3 Portfolio pada Drug Discovery

### 2.3.1 Aktivitas Molekul

Tingkat kesuksesan molekul menggambarkan hal yang sama dengan nilai *return* pada saham. Salah satu ukuran parameter untuk aktifitas molekul yang dapat digunakan adalah IC50. Semakin kecil nilai IC50 maka semakin baik molekul tersebut. karena hal ini berbandik terbalik dengan prinsip probabilitas kesuksesan. Nilai aktivitas molekul menjadi faktor untuk pengukuran nilai probabilitas kesuksesan. Untuk mendapatkan nilai aktivitas untuk setiap molekul digunakan persamaan sebagai berikut : [2]

$$a_i = \frac{1}{IC_i} \quad (2-1)$$

Keterangan :

$a_i$  = nilai aktivitas molekul i yang diberikan

$IC_i$  = Nilai aktivitas molekul i berdasarkan parameter IC50

### 2.3.2 Probabilitas Kesuksesan

Probabilitas Kesuksesan merupakan suatu parameter yang dijadikan acuan untuk melihat peluang keberhasilan molekul untuk menjadi kandidat portfolio. Nilai tersebut dapat diperoleh dengan persamaan sebagai berikut : [2]

$$p_i = a_i \cdot \bar{p} \cdot \frac{nTotal}{\sum_{i=1}^{nTotal} a_i} = \frac{a_i}{\sum_{i=1}^{nTotal} a_i} \quad (2-2)$$

Keterangan :

$p_i$  = nilai probabilitas kesuksesan untuk molekul i

$a_i$  = nilai aktivitas untuk molekul i,

$\bar{p}$  = nilai rata-rata probabilitas dari 1 individu

nTotal = Jumlah molekul dalam 1 individu

### 2.3.3 Expected Return

Expected Return merupakan nilai pengembalian yang diharapkan berdasarkan komposisi yang sudah ditetapkan dalam sebuah portfolio. Dengan menggunakan nilai yang didapatkan pada persamaan (2-5) dapat dihitung nilai Expected return, yaitu [2] :

$$E(X) = \frac{\sum_{i=1}^{nTotal} p_i}{nTotal} \quad (2-3)$$

Keterangan :

$E(X)$  = Nilai expected return untuk individu X

$p_i$  = nilai probabilitas kesuksesan untuk molekul i

### 2.3.4 Molecular Fingerprint

Molecular Fingerprint menggambarkan struktur dari suatu molekul yang dapat dijadikan acuan untuk membedakan antar struktur molekul. Fingerprint ini dapat direpresentasikan menjadi sebuah bit String [6] .

### 2.3.5 Diversity

Diversity merupakan nilai yang merepresentasikan keberagaman suatu populasi. Semakin besar nilai diversity, maka semakin besar pula tingkat keragaman suatu populasi. Nilai diversity dapat ditentukan dari nilai similarity untuk suatu individu, dimana nilai similarity dapat dihitung dengan melihat molecular fingerprint dan persamaan Tanimoto Similarity, seperti berikut [6] :

$$Sim_T(A, B) = \frac{N_{AB}}{N_A + N_B + N_{AB}} \quad (2-4)$$

Keterangan :

$N_A$  = Jumlah bit 1 pada molekul a saja

$N_B$  = Jumlah bit 1 pada molekul b saja

$N_{AB}$  = Jumlah bit 1 pada molekul a dan b

Sehingga nilai diversity dapat dihasilkan melalui persamaan:

$$d(x_i, x_j) = \frac{1}{Sim_T} \quad (2-5)$$

## 2.4 Non-Dominated Sorting Genetic Algorithm (NSGA-II)

### 2.4.1 Algoritma Genetika

Pada algoritma NSGA-II pemilihan parent ditentukan dengan melakukan pengurutan individu dengan mempertimbangkan nilai dominansi antar individunya (Non-Dominated Sorting) pada populasi. Individu yang mempunyai nilai rank tertinggi menggambarkan individu dengan nilai fitness terbaik. Selain melihat nilai fitness, algoritma ini juga memerhatikan nilai kedekatan antar individu dengan individu disampingnya (Crowding Distance). Individu dengan nilai crowding distance terbesar yang akan dipilih sebagai parent untuk menghasilkan individu baru dengan menggunakan crossover dan mutasi [7].

### 2.4.2 Non-Dominated Sorting

Non-Dominated Sorting merupakan suatu metode untuk mengurutkan individu pada suatu populasi dengan memerhatikan dominansi antar individu. Suatu individu p dikatakan mendominasi individu q jika salah satu atau kedua nilai fungsi objektif individu p lebih baik dari individu q. [8] [9].

### 2.4.3 Crowding Distance

crowding distance merepresentasikan jarak kerapatan antara front satu dengan front yang berada di dekatnya. Perhitungan ini dilakukan hanya pada tingkat front yang sama [10].

### 2.4.4 Crossover dan Mutation

Crossover merupakan proses persilangan antara 2 individu (parent) yang berbeda dan telah terpilih pada proses sebelumnya. Kedua individu ini nantinya akan disilangkan untuk menghasilkan individu baru (offspring). Setelah dilakukan crossover, langkah selanjutnya adalah melakukan mutasi. Mutasi dilakukan dengan cara mengubah beberapa gen secara acak pada individu yang dihasilkan dari crossover [8].

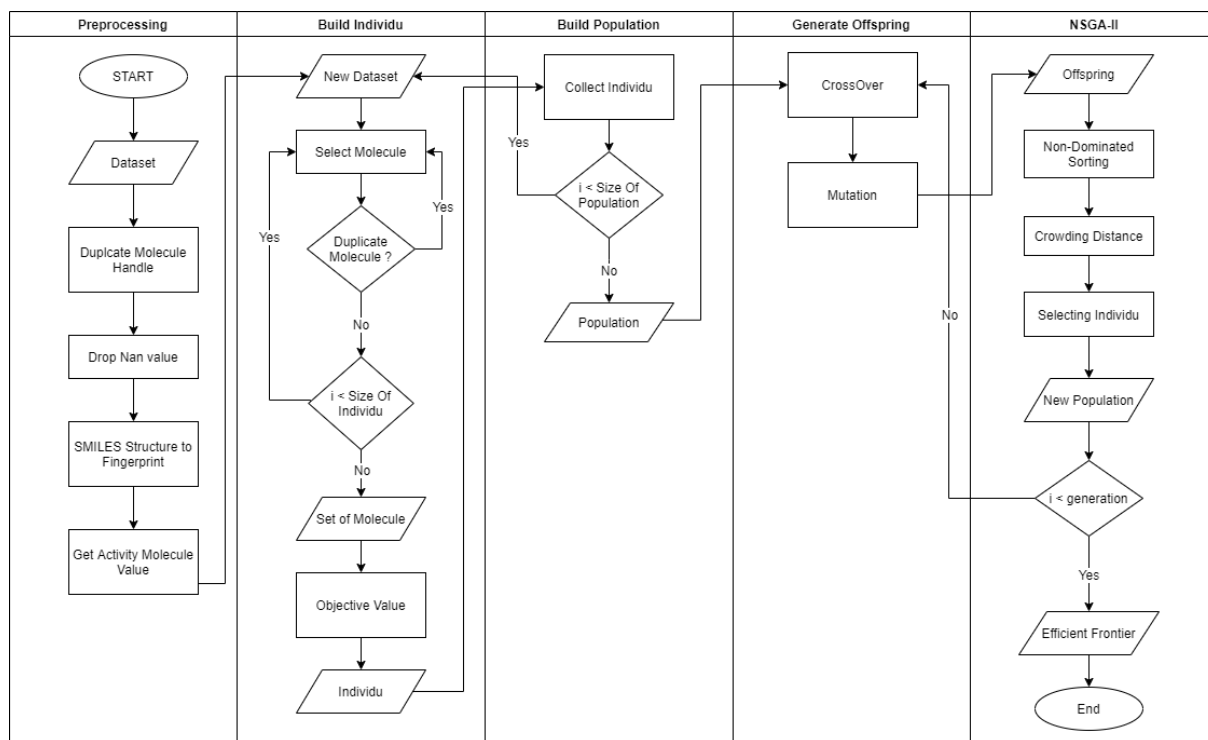
## 2.5 Penelitian Terkait

Dalam penelitiannya, Yevseyeva [2] menyimpulkan bahwa Portfolio pada keuangan dapat digunakan dalam masalah *drug discovery*. Algoritma genetika yang digunakan yaitu NSGA-II dan SMS-EMOA. Hasil yang diperoleh menunjukkan SMS-EMOA menawarkan nilai return yang lebih tinggi dibandingkan dengan NSGA-II. Hal ini dikarenakan pada NSGA-II diperlukan untuk mencari nilai inverse matrix, dimana dalam praktiknya sulit.

Menurut penelitian yang dilakukan oleh Anagnostopoulos [11] dalam membandingkan algoritma PESA, NSGA-II dan SPEA2 mendapatkan hasil bahwa PESA mempunyai performansi lebih baik secara pemilihan kedekatan individu pada *efficient frontier*. Sedangkan algoritma NSGA-II dan SPEA2 memiliki rata rata performansi yang baik untuk *hypervolume indicator*. Kelebihan lainnya dari kedua algoritma ini adalah bersifat fleksibel untuk menentukan *constraint* dan objektif untuk risiko.

### 3. Sistem yang Dibangun

Sistem yang akan dibangun pada penyelesaian Tugas Akhir ini digambarkan dalam bentuk diagram alir pada Gambar 1 di bawah ini.



Gambar 1 Perancangan Sistem

#### 3.1 Data Praprocessing

Dataset yang digunakan untuk pembangunan sistem ini berasal dari [www.ebi.ac.uk](http://www.ebi.ac.uk). Sebelum dataset tersebut digunakan, praproses data perlu dilakukan. Beberapa preproses data yang dilakukan adalah sebagai berikut :

##### 3.1.1 Drop Nan Value

Beberapa molekul memiliki Nan value pada salah satu fiturnya. Dengan adanya Nan value ini mengakibatkan beberapa proses yang akan ada tidak dapat dilanjutkan. Untuk mengatasi hal ini dilakukan drop value pada molekul yang memiliki Nan value ini. Berikut merupakan beberapa molekul yang mempunyai Nan value.

Molecule Name	SMILES	Activity		
		Standard Relation	Standard Value	Standard Units
CHEMBL179166	Nan	Nan	Nan	Nan
CHEMBL179166	Nan	'='	152000	nM
CHEMBL86473	Nan	'='	308000	nM
CHEMBL86473	Nan	'>'	10000000	nM
CHEMBL179166	Nan	Nan	Nan	Nan
CHEMBL179166	Nan	'='	78300	nM
CHEMBL2000194	Nan	'='	79300	nM



### 3.1.2 Duplicate Molecule Handle

Pada dataset yang tersedia memiliki beberapa nama molekul yang sama namun memiliki nilai activity dan SMILES yang berbeda. Untuk mengatasi hal ini untuk nama data molekul yang sama akan di drop dari dataset.

**Tabel 1 Perubahan Molekul dengan Nama yang Sama**

Sebelum	Sesudah
CHEMBL89067	CHEMBL89067(1)
CHEMBL89067	CHEMBL89067(2)
CHEMBL89067	CHEMBL89067(3)
CHEMBL89067	CHEMBL89067(4)
CHEMBL89067	CHEMBL89067(5)
CHEMBL8767	CHEMBL8767(1)
CHEMBL8758	CHEMBL8758(1)
CHEMBL8758	CHEMBL8758(2)
CHEMBL8758	CHEMBL8758(3)
CHEMBL8758	CHEMBL8758(4)

### 3.1.3 SMILES to Fingerprint Structure

Nilai Similarity antar molekul didapatkan dengan menghitung jarak antar molekul dengan menggunakan struktur Fingerprint antar molekul. Pada dataset hanya disediakan struktur SMILES, sehingga perlu diubah dalam struktur Fingerprint. Berikut merupakan contoh perubahan struktur yang dilakukan pada dataset :

Nama Molekul	SMILES	Fingerprint
CHEMBL140954	<chem>O=P(O)(O)C(F)(F)c1ccc(COc2ccc(OCc3ccc(C(F)(F)P(=O)(O)O)cc3)cc2)cc1</chem>	[1 0 0 ... 0 0 0]
CHEMBL440955	<chem>O=C1C[C@@H](c2ccc(C[C@H](Nc3nc4ccccc4s3)c3nc4ccccc4[nH]3)cc2)S(=O)(=O)N1</chem>	[1 0 1 ... 1 0 1]
CHEMBL246869	<chem>O=C1CC(c2ccc(C[C@H](Nc3nc4ccc(Br)cc4s3)c3nc4ccccc4[nH]3)cc2)S(=O)(=O)N1</chem>	[1 0 1 ... 1 0 1]
CHEMBL210836	<chem>O=C1C[C@@H](c2ccc(C[C@H](NS(=O)(=O)c3ccccc3)c3nc4ccccc4[nH]3)cc2)S(=O)(=O)[N-]1.[Na+]</chem>	[1 0 1 ... 1 0 1]
CHEMBL382311	<chem>CC(=O)N[C@@H](Cc1ccccc1)C(=O)N[C@@H](Cc1ccc([C@@H]2CC(=O)NS2(=O)=O)cc1)C(N)=O</chem>	[0 1 1 ... 1 0 1]
CHEMBL1935500	<chem>C/C=C/C=C/C(=O)N1Cc2cc(OCCc3nc/C=C/C(C)CCC)oc3C)ccc2C[C@H]1C(=O)O.CC(C)(C)N</chem>	[1 1 1 ... 1 0 1]
CHEMBL1935493	<chem>C/C=C/C=C/C(=O)N1Cc2cc(OCCc3nc/C=C/C(C)oc3C)ccc2C[C@H]1C(=O)O.CC(C)(C)N</chem>	[1 1 1 ... 1 0 1]
CHEMBL253500	<chem>O=C1CC(c2ccc(CC(NS(=O)(=O)c3ccccc(F)c3)c3ncc(CCCc4ccccc4)[nH]3)cc2)S(=O)(=O)N1</chem>	[1 0 1 ... 1 0 1]

### 3.1.4 Get Activity Molecule Value

Dalam pembangunan sistem ini molekul yang digunakan hanya yang memenuhi beberapa syarat, seperti :

1. Relasi antara molekul dan nilai aktivitasnya yang digunakan adalah '='
2. Satuan unit yang digunakan adalah nM
3. Nilai aktivitas molekul kurang dari 10000 nM

Setelah beberapa molekul terpilih, dilakukan konversi dari satuan nM ke mM. Berikut merupakan beberapa contoh molekul yang terpilih :

Molekul dari Dataset				Molekul terpilih			
Molecule Name	Standard Relation	Standard Value	Standard Units	Molecule Name	Standard Relation	Standard Value	Standard Units
CHEMBL440955	'='	2700	nM	CHEMBL440955	'='	2.7	mM
CHEMBL246869	'='	5000	nM	CHEMBL246869	'='	5	mM
CHEMBL381633	'>'	1000000	nM	CHEMBL210836	'='	0.32	mM
CHEMBL1946252	'>'	100000000	nM	CHEMBL382311	'='	2.4	mM
CHEMBL210836	'='	320	nM				
CHEMBL382311	'='	2400	nM				
CHEMBL2414203	'>='	200	ug.mL-1				
CHEMBL370470	'='	3801893963	nM				

### 3.2 Build Individu

Suatu individu terdiri dari beberapa molekul unik yang memiliki nilai activity dan SMILES nya masing-masing. Setelah proses pemilihan molekul dilakukan, nilai Activity dan Diversity dapat dicari. Kedua nilai ini yang nantinya akan digunakan untuk melakukan evaluasi untuk membentuk setiap individu dalam populasi setiap generasinya berdasarkan algoritma NSGA-II. Pemilihan molekul untuk membentuk individu pada generasi pertama dilakukan secara random.

### 3.3 Build Population

Beberapa individu yang telah dibentuk akan dikumpulkan menjadi satu populasi. Pada sistem ini, tidak dimungkinkan terdapat beberapa individu yang memiliki himpunan molekul yang sama persis antar molekul tersebut. Proses pembentukan individu akan terus dilakukan sampai batas jumlah individu dalam populasi terpenuhi.

### 3.4 Generate Offspring

Pada proses ini jumlah populasi ditambahkan menjadi 2 kali lipat dari jumlah populasi awal. Populasi baru ini terdiri dari parent atau himpunan individu dari populasi sebelumnya dan offspring yang akan dibentuk dari parent dengan beberapa proses sebagai berikut :

#### 3.4.1 Crossover

Crossover merupakan proses pertukaran silang yang dilakukan antara 2 individu (parent) dan menghasilkan 2 individu baru (offspring). Individu dipilih berdasarkan Probabilitas Crossover (PCo) yang telah ditentukan. Setiap iterasi dalam pemilihan individu dilakukan generate bilangan antara 0 sampai 1 secara random. Untuk setiap individu yang memiliki nilai probabilitas di atas nilai (PCo) akan dilakukan proses crossover.

#### 3.4.2 Mutation

Mutation merupakan proses pergantian beberapa molekul terpilih yang ada pada individu dengan molekul yang ada di dataset. Pemilihan molekul dilakukan dengan men-generate bilangan random antara 0 sampai 1 untuk setiap iterasi molekul. Nilai yang memiliki nilai kurang dari nilai Probabilitas Mutation (PMut) yang telah diberikan.

### 3.5 NSGA-II

NSGA-II merupakan salah satu algoritma yang dapat digunakan untuk menyelesaikan masalah Multi Objective Optimization (MOO). Algoritma ini salah satu jenis Evolutionary Algorithm (EA). Algoritma ini akan menghasilkan beberapa individu terbaik. Offspring yang telah dibentuk pada proses Generate Offspring akan dilakukan beberapa proses pemilihan individu terbaik untuk setiap generasinya. Iterasi akan terus berlangsung sampai jumlah generasi terpenuhi. Berikut merupakan beberapa proses yang dilakukan pada algoritma ini :

#### 3.5.1 Non-Dominated Sorting

Seperti yang sudah dijelaskan pada bagian .. Berikut merupakan langkah yang harus dilakukan pada Non-Dominated Sorting , yaitu : [8] [9].

1. Untuk setiap individu p yang berada pada populasi A , dilakukan :

- a. Inisialisasi  $S_p = \{ \}$ , dimana  $S_p$  merupakan himpunan individu  $p$  yang mendominasi  $q$ .
- b. Inisialisasi  $n_p = 0$ , dimana  $n_p$  menyimpan jumlah individu  $q$  yang mendominasi  $p$ .
- c. Untuk setiap individu  $q$  dalam populasi  $A$  akan diperiksa :
  - i. Jika  $p$  mendominasi  $q$ , maka tambahkan  $q$  kedalam himpunan  $S_p$  atau dapat ditulis sebagai  $S_p = S_p \cup \{q\}$
  - ii. Jika  $q$  mendominasi  $p$ , maka nilai  $n_p$  ditambahkan 1 atau dapat ditulis sebagai  $n_p = n_p + 1$
- d. Jika tidak ada satupun individu  $q$  yang mendominasi  $p$  ( $n_p = 0$ ), maka lakukan :
  - i. Individu  $p$  menjadi front pertama.
  - ii. Individu  $p$  diberi rank 1 atau dapat ditulis sebagai  $prank = 1$
  - iii. Update front pertama dengan menambahkan individu  $p$  ( $F1 = F1 \cup \{p\}$ )
2. Inisialisasi front = 1 dan  $i = 1$
3. Jika front ke  $i$  mempunyai anggota didalamnya ( $F_i \neq \{ \}$ ), maka lakukan :
  - a. Inisialisasi  $Q = \{ \}$ , dimana  $Q$  merupakan himpunan individu untuk front setelahnya ( $F_{i+1}$ )
  - b. Untuk setiap individu  $p$  di dalam  $F_i$ , dilakukan :
    - i. Untuk setiap individu  $q$  yang merupakan anggota  $S_p$ , dilakukan :
      - 1) Lakukan pengurangan jumlah untuk himpunan individu yang mendominasi  $q$  atau dapat ditulis sebagai  $n_q = n_q - 1$ , dimana  $n_q$  merupakan jumlah individu yang mendominasi  $q$ .
      - 2) Jika tidak ada satupun individu yang mendominasi  $q$  atau dapat ditulis sebagai  $n_q = 0$ , maka lakukan :
        - a. Individu  $q$  diberi rank  $i+1$  ( $qrank = i + 1$ )
        - b. Update  $Q$  dengan menambahkan individu  $q$  ( $Q = Q \cup q$ )
        - c. Tetapkan  $Q$  yang sudah terbentuk sebagai front berikutnya ( $F_i = Q$ )

### 3.5.2 Crowding Distance

Adapun cara mendapatkan nilai crowding distance yaitu [10]:

1. Inisialisasi distance untuk semua individu ke-  $k$  sampai  $n$  dengan nilai 0, dengan  $n$  merupakan jumlah individu yang ada pada satu front.
2. Untuk setiap fungsi objektif  $m$  dilakukan :
  - a. Pengurutan setiap individu dengan melihat nilai fungsi objektif  $m$ , sehingga  $I = \text{sort}(I, m)$
  - b. Inisialisasi  $I[1] = I[n] = \infty$
  - c. Untuk setiap individu ke  $k+1$  sampai  $n-1$  dilakukan :
    - i. Proses perhitungan jarak untuk masing – masing individu dengan tetangga terdekatnya menggunakan rumus :

$$I[k] = I[k] + \left( \frac{I[k+1].m - I[k-1].m}{f_m^{\max} - f_m^{\min}} \right)$$

Keterangan :

- $I[k].m$  = nilai fungsi objektif  $m$  dari individu  $k$  di  $I$   
 $f_m^{\max}$  = nilai maksimum fungsi objektif  $m$   
 $f_m^{\min}$  = nilai minimum fungsi objektif  $m$

### 3.5.3 Selecting Individu

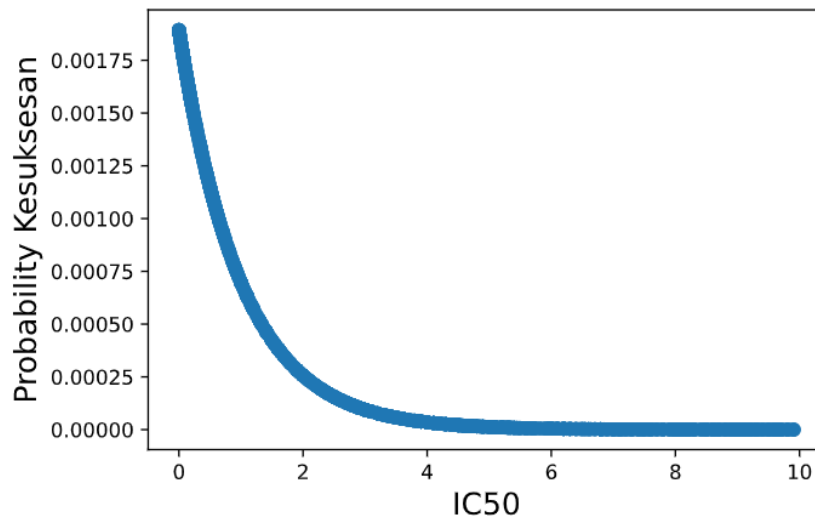
Nilai yang didapat pada Proses Crowding Distance digunakan pada proses Selecting Individu. Nilai dari setiap individu yang terdapat pada offspring tersebut dilakukan Ascending Order. Berdasarkan kumpulan molekul yang telah diurutkan, dipilih beberapa molekul pertama sebanyak besarnya populasi yang telah ditentukan. Kumpulan molekul terpilih ini yang akan dijadikan populasi baru untuk generasi selanjutnya.

## 4. Evaluasi

Dataset yang digunakan pada penelitian kali ini diambil dari [www.ebi.ac.uk](http://www.ebi.ac.uk). Dataset molekul yang digunakan merupakan molekul dengan target protein PTP1B. Pada kategori ini, data yang tersedia sebanyak 3715 molekul. Data tersebut kemudian dilakukan preprocessing untuk mendapatkan data yang dapat diolah. Setelah dilakukan preprocessing data, molekul yang dapat digunakan adalah sebanyak 1852 molekul. Data inilah yang kemudian digunakan untuk melakukan pencarian individu terbaik.

#### 4.1 Hasil Pengujian

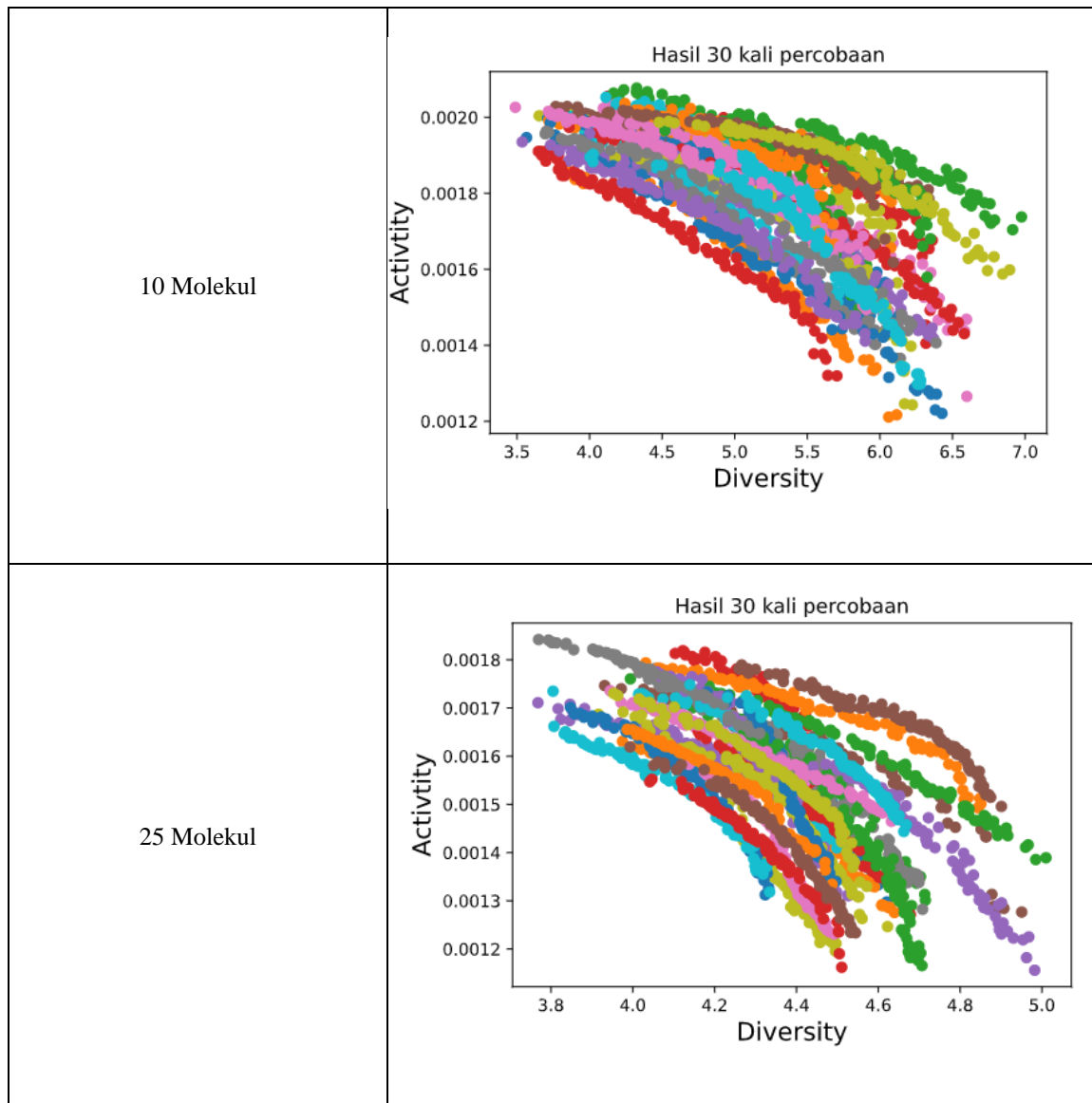
Berikut merupakan persebaran hubungan data aktivitas molekul dengan probailitas kesuksesan suatu molekul, seperti berikut :



Berdasarkan gambar ... dapat dilihat bahwa semakin besar nilai IC50 untuk setiap molekul maka tingkat probabilitas kesuksesan untuk molekul tersebut juga semakin kecil.

Tabel .. berikut merupakan hasil dari 30 kali percobaan running terhadap program dengan beberapa parameter yang digunakan seperti probabilitas Crossover (PCo) sebesar 0.5, Probabilitas Mutation sebesar 0.2, dan maximum generasi sebanyak 50 generasi. Adapun parameter jumlah molekul yang diubah untuk dilihat perbandingan dari masing-masing hasilnya. Hasil yang dimaksud adalah grafik pareto untuk masing-masing pengujian.

Jumlah Molekul	Grafik
5 Molekul	<p>Hasil 30 kali percobaan</p> <p>A scatter plot titled 'Hasil 30 kali percobaan' showing Activity (y-axis) versus Diversity (x-axis). The y-axis ranges from 0.0008 to 0.0020 with major ticks every 0.0002 units. The x-axis ranges from 3 to 9 with major ticks every 1 unit. The plot contains numerous data points of various colors (blue, green, red, purple, orange, etc.), forming a dense cloud that generally trends downwards from left to right, indicating a negative correlation between Activity and Diversity.</p>



#### 4.2 Analisis Hasil Pengujian

Berdasarkan hasil pengujian terlihat bahwa untuk ketiga hasil pengujian didapatkan grafik yang tidak konvergen. Hal ini menandakan bahwa program belum mencapai performansi yang optimal untuk menghasilkan himpunan individu terbaik. Analisa sementara penyebabnya adalah nilai generasi yang belum mencapai tingkat terbaik dan persebaran data untuk aktivitas molekul yang tidak seimbang. Sampai tulisan ini dimuat, penulis masih terus memperbaiki program tersebut.

### 5. Kesimpulan

Dari pembahasan teori dan beberapa pengujian yang telah dilakukan, maka dapat diambil kesimpulan bahwa metode pemilihan saham untuk portfolio keuangan dapat diterapkan pada kasus pemilihan molekul untuk portfolio molekul penyusun obat. Algoritma NSGA-II dapat menghasilkan beberapa individu terbaik dengan probabilitas kesuksesan maksimum untuk tingkat diversity tertentu. Ketiga hasil pengujian menghasilkan grafik dengan tingkat konvergensi yang rendah, hal ini menandakan bahwa performansi program belum mencapai batas optimal.

## Daftar Pustaka

- [1] J. Sun, C. Qu, Y. Wang, H. Huang, M. Zhang, H. Li, Y. Zhang, Y. Wang and W. Zou, "PTP1B, A Potential Target of Type 2 Diabetes Mellitus," *Molecular Biology*, vol. 5, 2016.
- [2] I. Yevseyeva, E. B. Lenselink, A. d. Vries, A. P. I. A. H. Deutz and M. T. Emmerich, " Application of portfolio optimization to drug discovery," *Information Sciences*, 2018.
- [3] I. Yunita, "Markowitz Model dalam Pembentukan Portfolio Optimal (Studi Kasus pada Jakarta Islamic Index)," *Jurnal Manajemen Indonesia*, 2018.
- [4] B. Pardosi and A. Wijayanto, "Aanalisis Perbedaan Return dan Risiko Saham Portfolio Optimal dengan Bukan Portfolio Optimal," *Mangement Analysis Journal*, 2015.
- [5] J. Andreas, "Introduction of Portfolio Risk," *Pinnacle Investment Research*, 2016.
- [6] I. Muegge and P. Mukherjee, "An overview of molecular fingerprint similarity search in virtual screening," *Expert Opinion on Drug Discovery*, pp. 137-148, 2015.
- [7] P. L. L. Belluano, "Optimalisasi Solusi Terbaik dengan Penerapan Non-Dominated Sorting II Algorithm," *Jurnal Ilmiah ILKOM*, 2016.
- [8] S. Isnaeni, D. Saepudin and R. F. Umbara, "Penerapan Algoritma Genetika Multi-objective NSGA-II Pada Optimasi Portofolio," *e-Proceeding of Engineering*, vol. 2, p. 6841, 2015.
- [9] A. Pratiwi, D. Saepudin and R. F. Umbara, "Optimasi Portofolio Mean-semivariance dengan Algoritma Genetika Multiobjective Evolutionary NSGA II," *e-Proceeding of Engineering*, vol. 5, pp. 8269-8281, 2015.
- [10] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," *IEEE Transaction on Evolutionary Computation*, vol. 6, pp. 182-197, 2002.
- [11] K. P. Anagnostopoulos and G. Mamanis, "Multiobjective evolutionary algorithms for complex portfolio optimization problems," *Springer*, pp. 259-279, 2019.