

**Implementasi Optimasi Portfolio
untuk Pemilihan Kandidat Molekul
dalam Menghambat Pertumbuhan PTP1B
pada Penyakit Diabetes Melitus**

Proposal Tugas Akhir

Kelas MK Penulisan Proposal (CCH4A3)

1301160792

Rafanzhani Elfarizy



**Program Studi Sarjana Informatika
Fakultas Informatika
Universitas Telkom
Bandung
2019**

Lembar Persetujuan

Implementasi Optimasi Portfolio untuk Pemilihan Kandidat Molekul dalam Menghambat Pertumbuhan PTP1B pada Penyakit Diabetes Melitus

*Implementation of Portfolio Optimization for Selection of Molecular
Candidates in Inhibiting PTP1B Growth in Diabetes Mellitus*

NIM :1301160792

Rafanzhani Elfarizy

Proposal ini diajukan sebagai usulan pembuatan tugas akhir pada
Program Studi Sarjana Informatika
Fakultas Informatika Universitas Telkom

Bandung, 15 November 2019

Menyetujui

Calon Pembimbing 1

Calon Pembimbing 2

Dr. DENI SAEPUKIN, S.Si., M.Si.

99750013

ISMAN KURNIAWAN, S.Pd., M.Si., M.Sc., Ph.D.

15870066

ABSTRAK

PTP1B atau *Protein Tyrosine Phosphatase 1B* merupakan protein yang menjadi salah satu penyebab dari penyakit Diabetes Melitus. Cara untuk menanganinya adalah dengan menghambat pertumbuhan dari 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. Pada tugas akhir ini, akan dilakukan pemilihan molekul untuk menghambat pertumbuhan PTP1B berdasarkan biaya yang disediakan dengan menggunakan metode optimasi portfolio pada keuangan. Terdapat 2 buah fungsi objektif yang harus dipertimbangkan, yaitu *expected return* yang direpresentasikan sebagai probabilitas kesuksesan dan *diversity* sebagai tingkat keberagaman antar molekul. kedua fungsi tersebut harus dimaksimalkan secara simultan. Permasalahan ini akan diselesaikan menggunakan *multi-objective optimization* dengan NSGA-II. Hasil akhir dari pemilihan ini berupa kumpulan solusi yang berisi kandidat molekul dengan tingkat kesuksesan tertinggi untuk nilai diversity tertentu.

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

Daftar Isi

Lembar Persetujuan	1
ABSTRAK	2
Daftar Gambar	4
Daftar Tabel	5
1. PENDAHULUAN	6
1.1. Latar Belakang	6
1.2. Perumusan Masalah.....	7
1.3. Tujuan	7
1.4. Rencana Kegiatan	7
1.5. Jadwal Kegiatan	8
2. KAJIAN PUSTAKA	9
2.1. <i>Drug discovery</i>	9
2.2. Manajemen Portfolio	9
2.3. Portfolio pada <i>Drug Discovery</i>	10
2.4. <i>Non-Dominated Sorting Genetic Algortihm (NSGA-II)</i>	15
2.5. Penelitian Terkait	20
3. PERANCANGAN SISTEM.....	22
3.1. Gambaran Umum Sistem	22
3.2. Rancangan Sistem	22
Daftar Pustaka	27

Daftar Gambar

Gambar 2.1 contoh dominance test.....	17
Gambar 3.1 <i>Flow-Chart</i> rancangan system utama.....	22
Gambar 3.2 <i>Flow-Chart</i> untuk Penerapan Algoritma NSGA-II.....	25

Daftar Tabel

Tabel 1.1 Jadwal Kegiatan	8
Tabel 2.1 Penelitian terkait	20

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*.

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. Perumusan Masalah

Berdasarkan latar belakang, dapat ditentukan permasalahan yang harus diselesaikan adalah sebagai berikut :

1. Bagaimana cara menentukan komposisi kandidat molekul penghambat pertumbuhan PTP1B dengan menggunakan metode optimasi portfolio pada keuangan ?
2. Bagaimana penerapan algoritma NSGA-II dalam menyelesaikan permasalahan *multi-objective optimization* ?

1.3. Tujuan

Berdasarkan perumusan masalah, tujuan yang ingin dicapai adalah sebagai berikut :

1. Menghasilkan komposisi kandidat molekul penghambat pertumbuhan PTP1B dengan nilai *expected return* tertinggi untuk suatu tingkat *diversity* tertentu dengan menggunakan metode optimasi portfolio keuangan
2. Mengimplementasikan algoritma NSGA-II untuk menyelesaikan permasalahan *multi-objective optimization*

1.4. Rencana Kegiatan

Rencana kegiatan yang akan dilakukan pada saat pengerjaan Tugas Akhir adalah sebagai berikut :

1. Pengumpulan Data

Pada tahap ini akan dilakukan pengumpulan data. Data yang akan diambil sebagai bahan untuk melakukan proses selanjutnya didapat dari www.ebi.ac.uk

2. Pra-Processing Data

Pra-processing data merupakan sebuah proses pemilihan data yang sesuai untuk dilakukan *processing data*. Data yang akan dihasilkan dari tahap ini adalah data yang siap untuk dilakukan *processing data*

3. *Processing Data*

Pada tahap ini akan dilakukan proses data yang sudah siap menggunakan metode *Solow-Polasky Diversity* sehingga menghasilkan nilai komposisi untuk setiap molekul yang ada

4. Analisa

Tahap terakhir yang dilakukan adalah menganalisis dari hasil yang didapatkan. Membandingkan dengan hasil penelitian yang sudah ada.

1.5. Jadwal Kegiatan

Tabel 1.1 Jadwal Kegiatan

Kegiatan	Bulan				
	Januari	Februari	Maret	April	Mei
Pengumpulan Data					
Pra-Processing Data					
Processing Data					
Analisa					

2. KAJIAN PUSTAKA

2.1. *Drug discovery*

2.1.1. *PTP1B*

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]

2.1.2. *Diabetes Melitus*

Diabetes melitus merupakan sebuah penyakit yang menyerang sistem metabolisme karena kurangnya asupan insulin. Salah satu penyebabnya adalah menurunnya resistensi pada pengiriman insulin yang disebabkan oleh PTP1B. Untuk mengobati penyakit ini salah satu langkah yang diambil adalah dengan menghambat pertumbuhan dari PTP1B. [1]

2.1.3. *Konsep Drug Discovery*

Drug discovery merupakan suatu proses untuk mencari suatu molekul yang dapat menghambat pertumbuhan suatu organisme yang dapat menyebabkan suatu penyakit. Untuk mendapatkan molekul terbaik diperlukan beberapa macam pengujian di laboratorium sebelum digunakan. Untuk mengurangi kegagalan, dilakukan pemilihan kandidat molekul dengan tingkat keberhasilan tertinggi. Dalam proses pemilihan tersebut dibatasi dengan pendanaan yang ditetapkan. Telah dibuktikan bahwa masalah semacam ini mempunyai kemiripan dengan pemilihan saham untuk portfolio. Pemilihan komposisi yang tepat akan menghasilkan tingkat keberhasilan yang tinggi untuk molekul menjadi sebuah bahan pembuatan obat. [2]

2.2. *Manajemen Portfolio*

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. Prinsip sederhana dari investasi adalah ketika investor ingin mendapatkan nilai

pengembalian dalam jumlah yang besar, maka investor juga harus menanggung nilai risiko yang besar pula. [3] .

Pembahasan tentang teori portfolio sudah pernah dijelaskan oleh Markowitz. Markowitz membuktikan dalam teorinya bahwa untuk membentuk suatu portfolio yang optimal, diperlukan pembagian modal untuk beberapa sekuritas investasi. Prinsip yang dipakai oleh Markowitz adalah menurunkan nilai risiko sampai ke titik paling rendah, sehingga akan didapatkan portfolio yang optimal. Proses menurunkan nilai risiko ini disebut diservikasi [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. Jika saham mengalami kenaikan maka investor akan mendapatkan untung, sebaliknya jika saham mengalami penurunan investor akan mengalami kerugian (risiko). Untuk masalah portfolio, ada nilai *expected return* atas kombinasi dari berbagai macam sekuritas. 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. Untuk masalah portfolio untuk keuangan, bobot diisi dengan menggunakan nilai *real* untuk setiap bobot aset yang ketika dijumlahkan akan sama dengan 1. Namun, pada masalah ini nilai bobot diisi dengan nilai 0 atau 1. Dengan kata lain pemilihan molekul sebagai aset tidak bisa dipilih hanya sebagian. Pemilihan diputuskan dengan memilih molekul (bobot = 1) atau tidak (bobot = 0). Selain itu, pemilihan aset untuk portfolio juga melihat nilai *budget* yang dikeluarkan. *Budget* yang dikeluarkan tidak boleh melebihi nilai *budget* yang telah ditentukan. [2]

2.3. Portfolio pada Drug Discovery

2.3.1. Fungsi Objektif

Pada masalah *drug discovery*, untuk mendapatkan portfolio yang optimal fungsi yang dipertimbangkan adalah *expected return* dan *diversity* . Kedua fungsi tersebut menjadi fungsi objektif yang harus dimaksimalkan secara simultan. Karena

terdapat lebih dari 1 fungsi objektif, permasalahan ini dapat diselesaikan dengan *multi-objective optimization*. Pada portfolio keuangan *output* yang dihasilkan berupa sekumpulan saham dengan *return* tertinggi untuk suatu risiko tertentu. Untuk masalah *drug discovery*, portfolio akan berisi sekumpulan molekul dengan tingkat probabilitas tertinggi untuk suatu tingkat diversity tertentu dalam menghambat organisme penyebab suatu penyakit. [2]

2.3.2. Aktivitas Molekul

Molekul yang mempunyai peluang untuk menghambat pertumbuhan PTP1B menggantikan peran saham sebagai sekuritas pada portfolio keuangan. Tingkat kesuksesan molekul menggambarkan hal yang sama dengan nilai *return*. Nilai aktivitas molekul menjadi faktor untuk pengukuran nilai probabilitas kesuksesan. Semakin tinggi nilai aktivitas yang diberikan oleh suatu molekul, maka semakin besar pula tingkat probabilitas kesuksesan untuk molekul tersebut. Untuk mendapatkan nilai aktivitas untuk setiap molekul digunakan persamaan sebagai berikut [2] :

$$a_i = e^{(l_i)} \quad (2-1)$$

Keterangan :

a_i = nilai aktivitas molekul i yang diberikan

l_i = nilai aktivitas (log) molekul i yang disediakan

Banyak parameter untuk menentukan nilai bio-aktivitas untuk suatu molekul. salah satu parameter yang biasa digunakan adalah IC50. Secara sederhana, IC50 dapat diartikan sebagai jumlah molekul yang harus dipakai (nM) untuk menghambat 50% pertumbuhan bakteri. Sehingga dapat disimpulkan bahwa semakin kecil nilai IC50 suatu molekul, maka semakin baik molekul tersebut. Maka agar nilai a_i pada persamaan (2-1) tetap merepresentasikan nilai aktivitas yang dibutuhkan digunakan persamaan sebagai berikut [2]:

$$a_i = e^{-(IC_i)} \quad (2-2)$$

Keterangan :

IC_i = Nilai aktivitas molekul i berdasarkan parameter IC50

2.3.3. Probabilitas Kesuksesan

Probabilitas Kesuksesan merupakan suatu parameter yang dijadikan acuan untuk melihat peluang keberhasilan molekul untuk menjadi kandidat portfolio. Hasil penjumlahan semua nilai probabilitas molekul pada portfolio sama dengan 1. Dengan begitu, didapatkan persamaan untuk menghitung nilai probabilitas sebagai berikut [2]:

$$\sum_i^{nTotal} p_i = 1 \quad (2-3)$$

Keterangan :

$nTotal$ = jumlah molekul dalam satu portfolio

p_i = probabilitas kesuksesan untuk molekul i

Berdasarkan persamaan (2-3) didapat persamaan untuk menghitung nilai rata - rata probabilitas kesuksesan untuk suatu portfolio, yaitu :

$$\bar{p} = \frac{1}{nTotal} \quad (2-4)$$

Keterangan :

\bar{p} = Nilai rata-rata probabilitas

Nilai rata-rata probabilitas ini akan digunakan untuk menentukan nilai probabilitas kesuksesan untuk masing-masing molekul. Nilai tersebut dapat diperoleh dengan persamaan sebagai berikut :

$$p_i = a_i \cdot \bar{p} \cdot \frac{nTotal}{\sum_{i=1}^{nTotal} a_i} = \frac{a_i}{\sum_{i=1}^{nTotal} a_i} \quad (2-5)$$

Keterangan :

p_i = nilai probabilitas kesuksesan untuk molekul i

a_i = nilai aktivitas untuk molekul I, diperoleh pada persamaan (2-2)

\bar{p} = nilai rata-rata probabilitas, diperoleh pada persamaan (2-4)

2.3.4. Expected Return

Expected Return merupakan nilai pengembalian yang diharapkan berdasarkan komposisi yang sudah ditetapkan dalam sebuah portfolio. Pada masalah *drug discovery*, nilai *expected return* dapat diartikan sebagai nilai pengembalian atas keberhasilan suatu molekul dalam menghambat perkembangan organisme penyebab penyakit. Untuk mendapatkan nilai *expected return* untuk suatu portfolio diperlukan nilai *expected return* untuk setiap molekul yang ada pada portfolio. Karena adanya ketidakpastian *return* untuk setiap molekul, maka nilai *expected return* setiap molekul tersebut dapat dimodelkan sebagai variable *random bernouli*. Sehingga, didapatkan persamaan sebagai berikut [2]:

$$\bar{x}_i = \begin{cases} \frac{G - c_i}{c_i}, & \text{kondisi ketika sukses} \\ -1, & \text{kondisi ketika gagal} \end{cases} \quad (2-6)$$

Keterangan:

\bar{x}_i = variabel *random bernouli* untuk molekul i

G = nilai *gain*, pada random variable bernouli dapat digantikan sebagai *budget* yang dialokasikan

c_i = harga molekul i

Nilai probabilitas kesuksesan yang diberikan adalah p_i untuk kondisi sukses dan $(1-p_i)$ ketika gagal. Berdasarkan persamaan (2-6) didapatkan persamaan untuk menghitung nilai *expected return* untuk setiap molekul I, yaitu:

$$E(\bar{x}_i) = \frac{G - c_i}{c_i} \cdot p_i - 1 \quad (2-7)$$

Keterangan:

$E(\bar{x}_i)$ = nilai *expected return* untuk molekul i

p_i = nilai probabilitas kesuksesan untuk molekul i, diperoleh pada persamaan (2-5)

Hasil dari penjumlahan nilai *expected return* untuk setiap molekul Berdasarkan persamaan (2-7) adalah nilai *expected return* untuk suatu portfolio. Sehingga, didapatkan persamaan untuk mencari nilai *expected return* untuk suatu portfolio, yaitu :

$$E(X) = \sum_{i=1}^{nTotal} E(\bar{x}_i) \quad (2-8)$$

Keterangan:

$E(X)$ = Nilai *expected return* untuk portfolio X

2.3.5. Solow-Polasky Diversity

Diversity merupakan nilai yang merepresentasikan keberagaman suatu populasi. Semakin besar nilai *diversity*, maka semakin besar pula tingkat keragaman suatu populasi. Nilai *diversity* ini diperlukan jika suatu kandidat molekul mengalami kegagalan, maka ada kemungkinan kandidat molekul lain yang berhasil, karena struktur pembentuk molekul yang berbeda. Solow dan Polasky telah menjelaskan teori mengenai *diversity* ini. Nilai *diversity* didapatkan dari *distance matrix* untuk setiap molekul yang tersedia. Adapun persamaan yang diberikan, sebagai berikut [2]:

Fungsi *distance* diperlukan untuk membentuk *distance matrix*, didapatkan dengan menggunakan *euclidean distance* dengan rumus :

$$d(x_i - x_j) = \sum_{k=1}^{nMol} \sqrt{(x_{ik} - x_{jk})} \quad (2-9)$$

Keterangan :

$d(x_i - x_j)$ = *Euclidean Distance Function*

x_i = Molekul ke-i

x_j = Molekul ke-j

x_{ik} = nilai bit ke-k struktur pembentuk molekul ke-i

x_{jk} = nilai bit ke-k struktur pembentuk molekul ke-j

Untuk nilai *bit vector* ditentukan oleh nilai keberadaan struktur dalam suatu molekul. kemudian akan dibentuk *covariance matrix* dengan rumus sebagai berikut :

$$F(x) = e^{-\theta d(x_i - x_j)} \quad (2-10)$$

Keterangan :

$F(x)$ = Covariance Matrix

θ = nilai konstanta yang ditentukan, pada masalah ini $\theta = 0,5$

$d(x_i - x_j)$ = Euclidean Distance, diperoleh dari persamaan (2-9)

Nilai diversity diinterpretasikan sebagai nilai jumlah yang didapat dari *inverse matrix* atas *covariance matrix* yang telah ditentukan sebelumnya. Sehingga, Persamaan *Solow-Polasky* dapat dituliskan sebagai :

$$D(x) = \sum_{i=1}^{nTotal} \sum_{j=1}^{nTotal} F(x)_{ij}^{-1} \quad (2-11)$$

Keterangan :

$D(x)$ = Nilai *diversity* untuk portfolio x

$nTotal$ = Jumlah aset yang ada di setiap portfolio

$F(x)_{ij}^{-1}$ = Nilai *diversity* untuk molekul i dengan j dari *inverse matrix* atas *covariance matrix*

2.4. Non-Dominated Sorting Genetic Algortihm (NSGA-II)

2.4.1. Algoritma Genetika

Algoritma genetika mengadopsi konsep evolusi yang dikembangkan oleh Darwin, dimana setiap generasi menghasilkan populasi yang lebih baik dari generasi sebelumnya. Populasi terdiri dari beberapa individu yang mempunyai beberapa gen penyusun. Individu terbaik dihasilkan dari hasil persilangan antara 2 individu (*parent*) yang menghasilkan individu baru (*offspring*). *Parent* dipilih berdasarkan nilai *fitnes* terbaik pada populasi. [6]

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. Nilai ini disebut sebagai *crowding distance*. Individu dengan nilai *crowding distance* terbesar lah yang akan dipilih

sebagai parent untuk menghasilkan individu baru dengan menggunakan *crossover* dan mutasi [6]

Pada tugas akhir ini individu digambarkan sebagai sebuah portfolio yang terdiri dari berbagai macam bobot untuk masing-masing molekul (gen). hasil akhir dari proses ini akan mengeluarkan beberapa solusi portfolio dengan *nilai expected return* tertinggi untuk suatu nilai *diversity* tertentu. [2]

2.4.2. Inisialisasi Populasi

Inisialisasi populasi merupakan tahap awal yang dilakukan pada algoritma NSGA-II. Inisialisasi dilakukan dengan membangkitkan secara acak sebuah populasi yang terdiri dari beberapa individu. Masing-masing individu terdiri dari beberapa gen yang berisi bobot untuk setiap aset. [7]

Pada tugas akhir ini, batasan untuk sebuah individu seperti dibahas pada bagian 2.2.3 dibatasi dengan 2 hal, yaitu :

$$\sum_{i=1}^{nGen} x_i = nTotal \quad (2-12)$$

$$\sum_{i=1}^{nTotal} x_i \cdot c_i = B \quad (2-13)$$

Keterangan :

$nGen$ = Jumlah gen dalam setiap individu

$nTotal$ = Jumlah individu dalam satu populasi

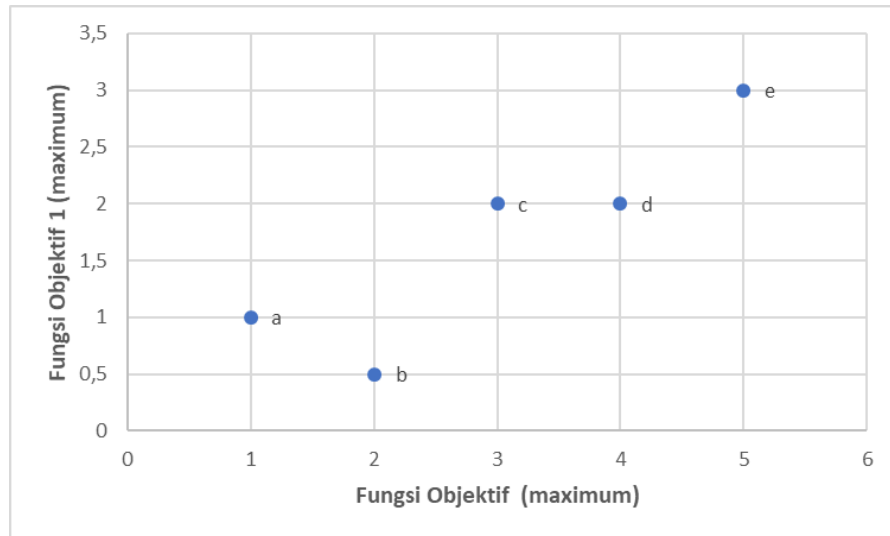
x_i = Bobot untuk molekul i

c_i = Harga untuk molekul i

B = Budget yang disediakan

2.4.3. Dominance Test

Dominance test merupakan sebuah pendekatan untuk memeriksa setiap individu yang mendominasi individu lainnya. Suatu individu p dikatakan mendominasi individu q jika salah satu atau kedua nilai fungsi objektif individu p lebih baik dari individu q [7].



Gambar 2.1 contoh dominance test

Berdasarkan Gambar 2.1 dapat dilihat bahwa :

- Antara c dan d didapatkan bahwa d mendominasi c, karena 2 nilai fungsi objektif d lebih baik dibanding c
- Antara c dan a didapatkan bahwa c mendominasi a, karena nilai fungsi objektif 1 untuk c lebih baik dibanding a
- Antara c dan e didapatkan bahwa c tidak mendominasi e, karena nilai kedua fungsi objektif c tidak lebih baik dari e

2.4.4. *Non-Dominated Sorting*

Non-Dominated Sorting merupakan suatu metode untuk mengurutkan individu pada suatu populasi dengan memerhatikan dominansi antar individu yang sudah dibahas pada bagian 2.4.2. Metode ini akan menghasilkan sebuah set solusi yang berisikan beberapa individu yang optimal. Berikut merupakan langkah yang harus dilakukan pada *Non-Dominated Sorting* , yaitu : [7].

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 merupakan variabel yang 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 atau dapat ditulis sebagai $n_p = 0$, maka lakukan :
 - i. Individu p menjadi *front* pertama.
 - ii. Individu p diberi *rank* 1 atau dapat ditulis sebagai $p_{rank} = 1$
 - iii. *Update front* pertama dengan menambahkan individu p atau dapat ditulis sebagai $F_1 = F_1 \cup \{p\}$
- 2. Inisialisasi $front = 1$ dan $i = 1$
- 3. Jika $front$ ke i mempunyai anggota didalamnya atau dapat ditulis sebagai $F_i \neq \{\}$, maka lakukan :
 - a. Inisialisasi $Q = \{\}$, dimana Q merupakan himpunan individu untuk *front* setelahnya atau dapat ditulis sebagai F_{i+1}
 - b. Untuk setiap individu p di dalam F_i , dilakukan :
 - i. Untuk setiap individu p 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* $i+1$ atau dapat ditulis sebagai $q_{rank} = i + 1$
 - b. *Update Q* dengan menambahkan individu q atau dapat ditulis sebagai $Q = Q \cup q$
 - c. Tetapkan Q yang sudah terbentuk sebagai *front* berikutnya atau dapat ditulis sebagai $F_i = Q$

2.4.5. 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. Adapun cara mendapatkan nilai *crowding distance* yaitu [8]:

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 *asscending* 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 :

$$I[k] = I[k] + \left(\frac{I[k + 1].m - I[k - 1].m}{f_m^{\max} - f_m^{\min}} \right) \quad (2-14)$$

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

2.4.6. Binary Tournament Selection

Binary Tournament Selection merupakan proses seleksi untuk semua individu. Suatu individu dikatakan lolos jika memenuhi 2 kriteria berikut [7] :

1. Non-domination Rank

Seleksi dilakukan dengan melihat *rank* yang telah didapat pada bagian sebelumnya. Nilai rank terkecil dipilih sebagai solusi.

2. Crowding Distance

Jika solusi yang memiliki nilai *rank* yang sama, maka solusi dengan nilai *crowding distance* terbesar yang akan dipilih menjadi solusi.

2.4.7. Crossover dan Mutasi

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*. [7]

Pada tugas akhir ini, portfolio dinyatakan sebagai *fixed-size* atau dalam kata lain bobot untuk masing-masing asset dinyatakan dengan 0 atau 1 (bukan bilangan *real*). Oleh karena itu untuk menentukan *offspring* yang akan dibentuk dengan melihat salah satu dari kedua *parent* yang dipilih secara random.

2.4.8. *Efficient frontier*

Setelah mendapatkan populasi dari suatu generasi akan dibangkitkan beberapa kemungkinan solusi optimal yang dapat diambil. Untuk mendapatkan portfolio yang optimal nilai *expected return* dan *diversity* harus dimaksimumkan secara simultan dengan memperhatikan *budget* yang diberikan. *Efficient frontier* digambarkan sebagai garis imajiner yang berisi sekumpulan individu yang memberikan nilai *expected return* tertinggi untuk suatu tingkat *diversity* tertentu. Untuk individu yang tidak berada pada garis ini dianggap sebagai solusi yang tidak optimal. [7]

2.5. Penelitian Terkait

Tabel 2.1 Penelitian terkait

No	Peneliti	Judul	Hasil
1	1. Iryna Yevseyeva 2. Eelke B. Lenselink 3. Alice de Vries 4. Adriaan P. IJzerman 5. Andr�e H. Deutz 6. Michael T.M. Emmerich	<i>Application of portfolio optimization to drug discovery</i>	Portfolio pada keuangan dapat digunakan dalam masalah <i>drug discovery</i> . 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 [2]
2	1. Sherly Isnaeni 2. Deni Saepudin 3. Rian Febrian Umbara	Penerapan Algoritma Genetika Multi-objective NSGA-II Pada Optimasi Portofolio Saham	Hasil yang didapatkan berupa [7]: 1. Dengan probabilitas diperkecil didapatkan hasil yang lebih baik (dilihat dari galat). 2. Penambahan jumlah saham menyebabkan

			<p>semakin sulit konvergensinya</p> <p>3. Hasil dari NSGA-II dinilai belum stabil, karena jumlah generasi yang belum cukup untuk mendapatkan hasil yang lebih baik</p>
3	<p>1. Anjar Pratiwi</p> <p>2. Deni Saepudin</p> <p>3. Rian Febrian Umbara</p>	<p>Optimasi Portofolio Mean-semivariance dengan Algoritma Genetika Multiobjective Evolutionary NSGA-II</p>	<p>Metode mean-semivariance menggunakan algoritma NSGA-II menghasilkan nilai yang lebih optimal dibandingkan dengan metode mean-variance (tanpa NSGA-II). Seluruh himpunan solusi pada <i>mean-semivariance</i> memiliki nilai return positif memiliki nilai resiko yang hampir mendekati 0. [9]</p>
4	<p>1. Konstantinos P. Anagnostopoulos</p> <p>2. Georgios Mamanis</p>	<p>Multiobjective evolutionary algorithms for complex portfolio optimization problems</p>	<p>Penelitian ini membandingkan algoritma PESA, NSGA-II dan SPEA2. Hasil yang didapatkan dari membandingkan 3 buah algoritma ini adalah PESA mempunyai perfomansi lebih baik secara pemilihan kedekatan individu pada <i>efficient frontier</i>. Sedangkan algoritma NSGA-II dan SPEA2 memiliki rata rata perfomansi yang baik untuk <i>hypervolume indicator</i>. Kelebihan lainnya dari kedua algoritma ini adalah bersifat fleksibel untuk menentukan <i>constraint</i> dan objektif untuk risiko [10]</p>

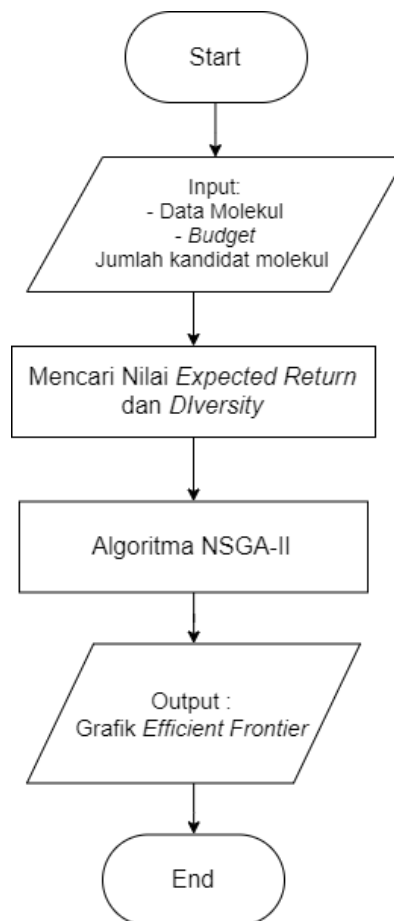
3. PERANCANGAN SISTEM

3.1. Gambaran Umum Sistem

Pada penyelesaian masalah untuk Tugas Akhir ini akan digunakan salah satu metode *multi-objective algorithm* yaitu NSGA-II. Secara umum, algoritma ini akan memilih individu terbaik yang berisikan gen-gen yang telah dipilih. Pemilihan gen ini didasari pada nilai dominansi menggunakan *non-dominance sorting*. Individu dengan *rank* terkecil merupakan solusi erbaik. Populasi yang dihasilkan dalam setiap generasi adalah kumpulan individu yang tidak saling mendominasi.

3.2. Rancangan Sistem

3.2.1. Sistem Utama



Gambar 3.1 *Flow-Chart* rancangan system utama

Dengan melihat Gambar 3.1, berikut merupakan rancangan system yang akan dibangun pada Tugas Akhir ini:

1. *Input* nilai yang dibutuhkan untuk membangun sistem ini. Adapun nilai yang dibutuhkan adalah :
 - a. Data Molekul merupakan data yang berisikan sekumpulan *record* molekul yang tersedia. Setiap *record* memiliki struktur molekul yang direpresentasikan sebagai bit vector dan nilai aktivitas. Data tersebut dapat diperoleh pada www.ebi.uk.ac yang merupakan layanan database penyediaan molekul.
 - b. *Budget* merupakan jumlah uang yang disediakan oleh *user* untuk membeli sejumlah molekul
 - c. Nilai jumlah kandidat molekul dalam suatu portfolio.
2. Menghitung nilai *Expected Return* dan *Diversity* untuk setiap portfolio yang terbentuk. Nilai ini kemudian akan digunakan untuk menentukan suatu portfolio optimal atau tidak.
3. Semua portfolio yang terbentuk akan dilakukan proses *sorting* dengan melihat nilai yang didapat pada poin (2) dengan menggunakan algoritma NSGA-II sesuai dengan metode *Markowitz with Solow-Polasky Diversity*.
4. *Output* yang dihasilkan dari Algoritma NSGA-II berupa grafik dari *efficient frontier*. Grafik ini menggambarkan sekumpulan portfolio yang memiliki nilai *Expected Return* tertinggi untuk setiap nilai *Diversity* tertinggi.

3.2.2. *Expected Return*

Berikut merupakan langkah yang harus dilakukan untuk mendapatkan nilai *Expected Return* untuk setiap portfolio yang mungkin, yaitu :

1. Berikut merupakan beberapa nilai variabel yang harus dimasukkan :
 - a. nilai Aktivitas (IC50) pada dataset untuk setiap molekul yang berada pada setiap portfolio
 - b. Nilai gain diperoleh dari jumlah dana yang dialokasikan (*Budget*).
 - c. Harga untuk setiap molekul, dimana nilai ini diberikan oleh dataset.

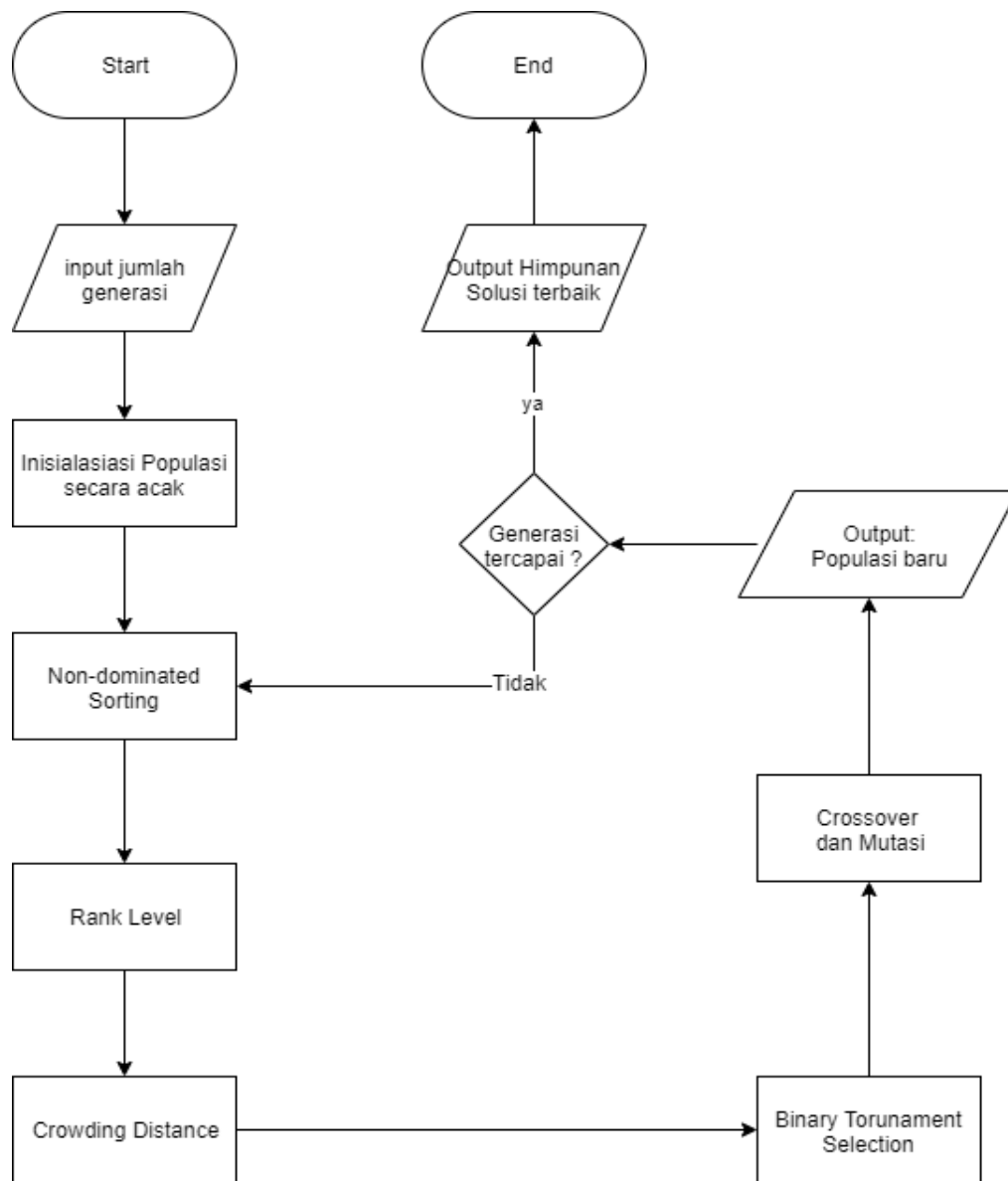
2. Melakukan proses penghitungan terhadap nilai aktivitas pada poin(1) dengan menggunakan persamaan (2-2). Sehingga mendapatkan nilai aktivitas untuk setiap molekul pada suatu portfolio.
3. Melakukan proses perhitungan dengan menggunakan persamaan (2-5) untuk mencari nilai probabilitas kesuksesan masing-masing molekul.
4. Melakukan proses perhitungan untuk mendapatkan nilai *expected return* dari nilai yang sudah didapat pada poin sebelumnya. Dengan menggunakan persamaan (2-7) akan didapatkan nilai *expected return* per molekulnya.
5. Nilai *expected return* untuk setiap portfolio merupakan hasil dari penjumlahan nilai *expected return* setiap molekul pada portfolio tersebut. dengan menggunakan persamaan (2-8) akan didapatkan nilai *expected return* untuk setiap portfolio yang dibentuk.
6. *Output* terakhir yang dihasilkan dari perhitungan *expected return* ini adalah nilai *expected return* untuk setiap portfolio yang akan menjadi pertimbangan untuk pemilihan portfolio yang optimal.

3.2.3. Nilai Diversity

Nilai diversity untuk suatu portfolio dapat diperoleh dengan melakukan beberapa langkah seperti berikut :

1. *Input* struktur molekul. struktur molekul direpresentasikan sebagai *bit vector* yang menyatakan struktur pembentuk molekul tersebut.
2. Melakukan proses perhitungan jarak antar molekul satu dengan yang lainnya. Dengan menggunakan persamaan (2-9) didapatkan nilai jarak untuk masing-masing molekul. Nilai-nilai ini akan dibentuk menjadi sebuah *distance matrix*
3. Membentuk *inverse matrix* yang terbentuk pada poin (2)
4. Mencari nilai diversity untuk setiap portfolio. Nilai ini dapat diperoleh dengan menggunakan persamaan (2-11)
5. *Output* yang dihasilkan berupa nilai *diversity* untuk setiap portfolio.

3.2.4. Algoritma NSGA-II



Gambar 3.2 Flow-Chart untuk Penerapan Algoritma NSGA-II

Berdasarkan Gambar 3.2 untuk menentukan beberapa portfolio terpilih dengan nilai diversity tertentu dan nilai *expected return* yang maksimal, diperlukan beberapa cara sebagai berikut :

1. *Input* jumlah generasi. Jumlah generasi menggambarkan banyaknya iterasi yang harus dilakukan untuk mendapatkan kumpulan portfolio yang optimal
2. Inisialisasi populasi dengan cara membangkitkan secara acak beberapa kemungkinan solusi dengan melihat batasan seperti yang sudah dibahas pada bagian 2.4.2.
3. Lakukan *sorting* dengan melihat dominansi antar individu seperti yang sudah dibahas pada bagian 2.4.4.
4. Setelah diurutkan tentukan nilai *fitness* dari rank yang sudah didapat pada poin (3). Individu dengan nilai *rank* terkecil mempunyai nilai *fitness* terbaik.
5. Jika ada individu dalam *front* yang sama, perlu dilakukan perhitungan nilai *crowding distance*. Perhitungan ini sudah dibahas pada bagian 2.4.5. Individu dengan nilai *crowding distance* terbesar yang akan terpilih.
6. Kemudian dilakukan pemilihan *parent* dengan mempertimbangkan nilai *rank* dan *crowding distance* seperti yang dijelaskan pada bagian 2.4.6.
7. Setelah *parent* terpilih dilakukan *crossover* dan mutasi untuk mendapatkan *offspring*. Proses ini sudah dijelaskan pada bagian 2.4.7.
8. Semua proses dilakukan sampai mencapai nilai generasi yang ditetapkan. Hasilnya berupa kumpulan solusi dengan individu terbaik.

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] P. L. L. Belluano, "Optimalisasi Solusi Terbaik dengan Penerapan Non-Dominated Sorting II Algorithm," *Jurnal Ilmiah ILKOM*, 2016.
- [7] 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.
- [8] 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.
- [9] A. Pratiwi, D. Saepudin and R. F. Umbara, "Optimasi Portofolio Mean-semivariance dengan Algoritma Genetika Multiobjective Evolutionary NSGA II," *e-Proceeding of Enginering*, vol. 5, pp. 8269-8281, 2015.
- [10] K. P. Anagnostopoulos and G. Mamanis, "Multiobjective evolutionary algorithms for complex portfolio optimization problems," *Springer*, pp. 259-279, 2019.