

Learning Struktur Bayesian Networks menggunakan Novel Modified Binary Differential Evolution pada Klasifikasi Data

Azmi Hafizha Rahman Z.A. ^{#1}, M. Syahrul Mubarak ^{#2}, Adiwijaya^{#3}

*School of Computing, Telkom University
Bandung, Indonesia*

¹azmi.hafizha@outlook.com

²msyahrulmubarak@telkomuniversity.ac.id

³adiwijaya@telkomuniversity.ac.id

Abstract

Bayesian Networks is one of the methods to model a probabilistical relationship discussed in Probabilistic Graphical Models. Bayesian Networks is constructed upon nodes and edges, the nodes is the representation of the random variables of the problem, and the edges is the representation of direct influence or dependency relationship between nodes. On a simple problem, the structure of a Bayesian Networks usually determined by an expert judgment or simply human intuition, problem arises when the problem is no longer simple and consists of so many nodes and edges combination possibilities. This research will try to construct an optimal Bayesian Networks structure using Novel Modified Binary Differential Evolution without the help from expert judgment or human intuition. Novel Modified Binary Differential Evolution is a binary discrete optimization problem algorithm with a binary chromosome representation, this algorithm is a further development of the standard Differential Evolution, because the former Differential Evolution is only able to work on a continuous optimization problem. When tested upon Alarm, Asia, Carpo, Insurance, and Water data set, the resulting BDeu score for each data set is -1973.77, -243.68, -2450.54, -2024.17, and -1621.90 respectively.

Keywords: Bayesian Networks, Structure Learning, Novel Modified Binary Differential Evolution, Differential Evolution

Abstrak

Bayesian Networks merupakan salah satu metode pemodelan probabilitas pada *Probabilistic Graphical Models*. *Bayesian Networks* terdiri dari *nodes* yang merepresentasikan variabel pada masalah yang dikaji dan *edges* yang merepresentasikan relasi dependensi antar *node*. Pada masalah yang sederhana, struktur *Bayesian Networks* biasanya ditentukan oleh ahli di bidang masalah tersebut atau berasal dari intuisi alami manusia. Perancangan struktur *Bayesian Networks* secara manual ini akan sulit dilakukan apabila kasus yang dikaji merupakan kasus yang kompleks yang memiliki sangat banyak *node* dan sangat banyak kemungkinan *edges* yang menghubungkannya. Pada penelitian ini, dilakukan pengujian dan analisa terhadap proses pencarian struktur *Bayesian Networks* menggunakan algoritma *Novel Modified Binary Differential Evolution*. *Novel Modified Binary Differential Evolution* merupakan algoritma optimasi permasalahan diskrit dengan representasi solusi berbentuk biner yang merupakan pengembangan dari algoritma *Differential Evolution*. Hasil pengujian terhadap data Alarm, Asia, Carpo, Insurance, dan Water masing-masing diperoleh skor BDeu sebesar -1973.77, -243.68, -2450.54, -2024.17, dan -1621.90.

Kata Kunci: Bayesian Networks, Perancangan Struktur, Novel Modified Binary Differential Evolution, Differential Evolution

I. PENDAHULUAN

Bayesian Networks (BN) merupakan salah satu metode reasoning yang memodelkan keterhubungan antar variabel pada *Probabilistic Graphical Models (PGM)*. Representasi utama pada BN adalah berupa *Directed Acyclic Graph (DAG)*, yang *nodenya* adalah variabel dan *edgenya* adalah penunjuk dependensi suatu *node* terhadap *node* yang lain [1]. BN banyak diminati oleh berbagai bidang dikarenakan kemampuannya yang dapat memetakan dependensi antar variabel, tidak seperti metode-metode *reasoning* yang lain. Namun untuk dapat memanfaatkan BN dengan efektif dan efisien, dibutuhkan struktur BN yang tepat. Pendekatan standar untuk membangun sebuah BN merupakan proses yang sangat memakan tenaga karena membutuhkan pengetahuan dari ahli bidang. Cara lain adalah membangun BN dari data yang dimiliki atau memperbaiki BN yang dibangun oleh ahli bidang dengan tujuan menambah tingkat efisiensi dan akurasi dari model yang dimaksud [2].

Novel Modified Binary Differential Evolution (NMBDE) merupakan metode baru berbasis *Differential Evolution (DE)* yang efisien untuk menyelesaikan permasalahan optimasi diskrit biner [3]. NMBDE mengenalkan operator baru yaitu *probability estimation operator* untuk menghasilkan individu anak, namun strategi-strategi asli dari DE standar dipertahankan agar tetap dapat memanfaatkan karakteristik DE seperti kemudahan pada implementasi dan kemudahan pada *tuning parameter* [3].

Penelitian ini berusaha menganalisis dan menyelesaikan masalah pencarian struktur BN menggunakan NMBDE serta menganalisis dan mengukur performansi struktur BN yang dihasilkan. Performansi NMBDE dibandingkan dengan performansi *Globally Optimal Bayesian Network learning using Integer Linear Programming (GOBNILP)* [4].

II. LANDASAN TEORI

A. Evaluasi Bayesian Networks

Untuk menilai kualitas struktur BN, dibutuhkan sebuah fungsi evaluasi yang nantinya nilai fungsi evaluasi ini akan menjadi fungsi fitness pada NMBDE. Ada banyak macam fungsi scoring yang dapat dijadikan acuan dalam menilai bagus atau tidaknya suatu BN. Salah satu fungsi evaluasi *Bayesian Networks* adalah *BDeu (Bayesian Dirichlet equivalence uniform) Score* [5]. Fungsi *BDeu* ini yang digunakan pada penelitian ini seperti terdefinisi pada Persamaan (1),

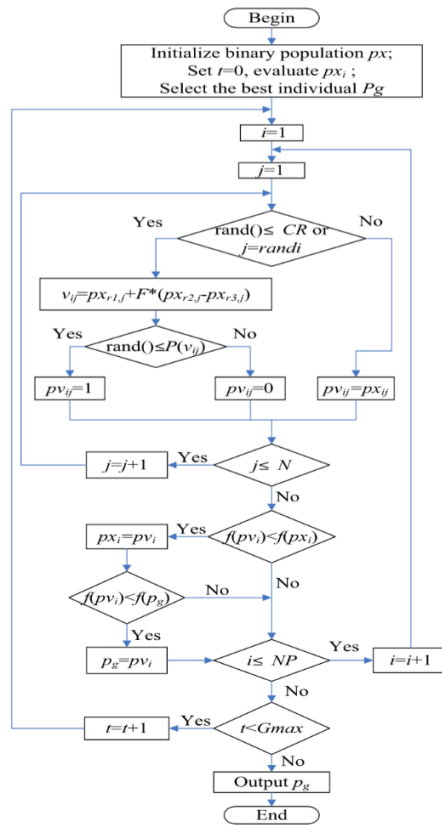
$$BDeu(N : D, \alpha) = \sum_i^n \sum_j^{q_i} \log \Gamma\left(\frac{\alpha}{q_i}\right) - \log \Gamma\left(\frac{\alpha}{q_i} + n_{ij}\right) + \sum_k^{r_i} \log \Gamma\left(\frac{\alpha}{r_i \cdot q_i} + n_{ijk}\right) - \log \Gamma\left(\frac{\alpha}{r_i \cdot q_i}\right) \quad (1)$$

dimana n merupakan jumlah *node* dalam hal ini adalah variabel, q_i merupakan jumlah instansiasi *parent node* i , r_i merupakan jumlah kemungkinan *value node* i dan α merupakan nilai parameter *ESS (Equivalent Sample Size)*.

Fungsi *BDeu* ini adalah fungsi maksimasi, yang artinya semakin tinggi nilai keluaran fungsi, maka semakin baik pula struktur BN yang sedang dievaluasi, indikator semakin baik disini berarti struktur BN yang dievaluasi semakin merepresentasikan hubungan dependensi antar variabel pada data yang diberikan, yang artinya nilai n_{ijk} pada iterasi k akan semakin besar dikarenakan semakin banyaknya baris data yang sesuai dengan tabel *CPD* yang terbentuk. Nilai maksimum yang mungkin dari suatu skor *BDeu* adalah 0.

B. Penjelasan NMBDE

NMBDE merupakan pengembangan dari *Differential Evolution* yang dikembangkan oleh Ling Wang, Xiping Fu, Yunfei Mao, Muhammad Ilyas Menhas, Minrui Fei dalam [3]



Gambar 1. Flowchart cara kerja algoritma NMBDE [3].

NMBDE mengadopsi skema representasi individu berupa string biner. NMBDE tetap menggunakan *updating formula DE* standar, termasuk diantaranya *operator* mutasi, *operator* rekombinasi dan *operator* penyeleksian. Namun karena *DE* standar mengoperasikan bilangan riil kontinu pada operasi mutasinya, NMBDE mengenalkan *probability estimation operator* agar proses mutasi dapat dilakukan terhadap bilangan diskrit biner [3]. Flowchart NMBDE digambarkan pada Gambar 1.

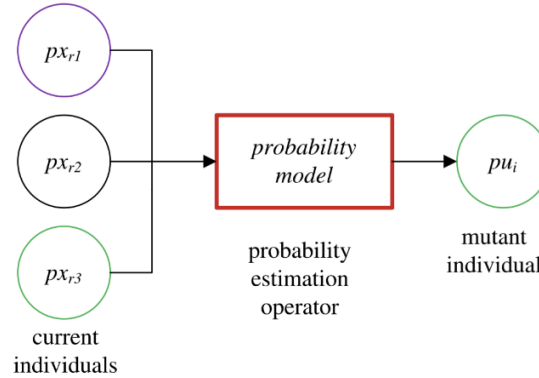
C. Probability Estimation Operator NMBDE

NMBDE mengenalkan operator baru pada algoritma *DE* yaitu *probability estimation operator (peo)* yang digunakan untuk membangkitkan individu baru hasil mutasi yang terencode dalam biner [3]. *Probability estimation operator* $P(px_{ij}^{t+1})$ didefinisikan pada Persamaan (2) sedangkan *Mutant Operator (MO)* yang terlibat didalamnya dijelaskan pada Persamaan (3),

$$P(px_{ij}^{t+1}) = 1/1[1 + e^{-2b * \frac{MO-0.5}{1+2F}}] \tag{2}$$

$$MO = px_{r1j}^t + F * (px_{r2j}^t - px_{r3j}^t) \tag{3}$$

dimana F adalah *scaling factor*; $px_{r1j}^t, px_{r2j}^t, px_{r3j}^t$ merupakan bit ke- j pada tiga buah individu yang dipilih secara acak; b merupakan *bandwidth factor* yang bernilai konstan bilangan riil positif. *Mutant operator* pada *DE* standar direpresentasikan dengan *MO*. Skema *probability estimation operator* dijelaskan pada Gambar 2.



Gambar 2. Skema peran peo pada pembentukan individu hasil mutasi NMBDE [3].

D. Mutasi NMBDE

Proses mutasi pada *NMBDE* memanfaatkan *probability estimation operator* untuk menentukan apakah gen pada suatu individu yang dipilih akan bernilai 0 atau 1. Formula translasi dari *peo* menjadi nilai pada gen dijelaskan pada Persamaan (4),

$$pu_{ij}^{t+1} = \begin{cases} 1, & \text{if } rand \leq P(px_{ij}^{t+1}) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

dimana *rand* adalah bilangan acak; $P(px_{ij}^{t+1})$ adalah komponen ke- j pada vektor *probability estimation operator* yang dimiliki individu ke- i [3].

E. Rekombinasi NMBDE

Operator rekombinasi digunakan untuk menghasilkan individu $pvi = (pvi1 , pvi2 , \dots , pviN)$ dengan menggabungkan individu *target* dan individu hasil mutasi dari target tersebut. Vektor hasil *pvi* akan dibentuk berdasarkan Persamaan (5) [3],

$$pv_{ij}^{t+1} = \begin{cases} pu_{ij}^{t+1}, & \text{if } (rand \leq CR) \text{ or } (j = r \text{ and } i) \\ px_{ij}^t, & \text{otherwise} \end{cases} \quad (5)$$

F. Seleksi Survivor NMBDE

Individu baru pv_i^{t+1} akan menggantikan individu lama px_i^t apabila memenuhi persamaan (6),

$$px_i^{t+1} = \begin{cases} pv_i^{t+1}, & \text{if } f(pv_i^{t+1}) < f(px_i^t) \\ px_i^t, & \text{otherwise} \end{cases} \quad (6)$$

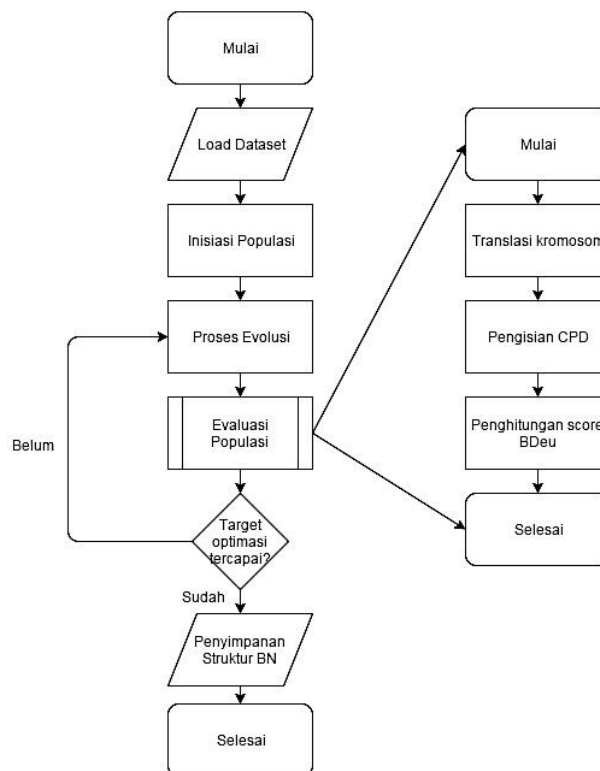
dimana fungsi f adalah fungsi *fitness* yang mengukur nilai *fitness* dari sebuah individu [3].

G. Flowchart NMBDE

Secara garis besar, alur *NMDBE* dapat digambarkan menjadi flowchart seperti pada Gambar 2. Flowchart *NMBDE* mirip dengan *DE* standar, letak perbedaannya adalah pada *encoding* kromosomnya, gen kromosom pada *NMBDE* berisikan nilai biner sedangkan *DE* standar berisikan nilai riil [3].

III. PERANCANGAN SISTEM

Tujuan dari penelitian ini adalah membuat sistem yang dapat menghasilkan struktur *BN* yang memiliki kualitas bagus dengan menggunakan algoritma *NMBDE* yang memiliki skor *BDeu* mendekati atau sama dengan skor *BDeu* hasil pencarian struktur menggunakan *GOBNILP*. Flowchart untuk sistem yang dibuat digambarkan pada Gambar 3. Proses Evaluasi Populasi memiliki sub proses yang memiliki keluaran akhir berupa skor *BDeu* untuk mengukur tingkat baik atau tidaknya struktur yang dijadikan masukan



Gambar 3. Flowchart aproksimasi struktur *BN* yang optimal menggunakan *NMBDE*.

A. Inisiasi Populasi

Sebuah populasi merupakan kumpulan dari kromosom-kromosom, sebuah kromosom merepresentasikan sebuah solusi untuk permasalahan yang ditangani, dalam hal ini adalah struktur *BN*. Jadi pada sistem ini sebuah kromosom menggambarkan sebuah struktur *BN*. Untuk detail penjelasan desain kromosom, akan dijelaskan pada bagian Evaluasi Kromosom.

Pada proses inisiasi populasi ini, terdapat dua buah parameter pengaturan yaitu jumlah kromosom pada populasi dan jumlah order bit (*ob*). Parameter jumlah kromosom berfungsi untuk mengatur seberapa ketat persaingan antar kromosom untuk generasi berikutnya, dan parameter *ob* yang mengatur urutan variabel pada matriks konektivitas pada translasi kromosom, parameter *ob* ini berpengaruh pada panjang kromosom yang dibangkitkan, panjang kromosom terdefinisi pada Persamaan (7),

$$\text{panjang kromosom} = (ob * vars) + (((vars * vars) - vars)/2) \tag{7}$$

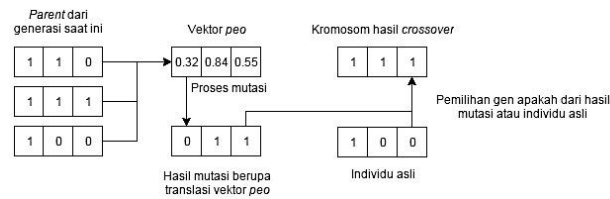
dimana *ob* adalah *order bit* sebagaimana yang telah dijelaskan sebelumnya, dan *vars* adalah jumlah variabel yang ada pada data set yang ingin dicari struktur *BN* nya, semakin banyak jumlah *order bit*, maka semakin kecil kemungkinan nilai *order bit* untuk sama.

B. Proses Evolusi

Proses evolusi terdiri dari proses modifikasi kromosom-kromosom pada populasi generasi saat ini, dengan tujuan memperbaiki kromosom-kromosom untuk generasi berikutnya.

Pada sistem ini, NMBDE diimplementasikan untuk melakukan proses evolusi, untuk detail penjelasan alur kerja NMBDE, dapat dilihat pada bagian *Flowchart NMBDE*.

Pada proses ini terdapat 3 buah *parameter* pengaturan yaitu *parameter scaling factor F*, *parameter bandwidth factor b*, dan *parameter* probabilitas dilakukannya mutasi lalu rekombinasi *CR*, ilustrasinya digambarkan pada Gambar 4.



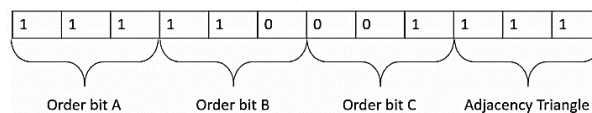
Gambar 4. Ilustrasi pembentukan kromosom hasil mutasi dan rekombinasi.

C. Evaluasi Populasi

Evaluasi populasi adalah proses penilaian masing-masing kromosom pada populasi di generasi yang dievaluasi terhadap skor *BDeu*.

D. Translasi Kromosom

Sebuah kromosom menggambarkan sebuah solusi untuk masalah yang ditangani. Solusi yang ingin dicari untuk permasalahan *structure learning* pada *BN* adalah berupa susunan *DAG* yang menggambarkan dependensi antar variabel random. Pada sistem yang dirancang, susunan *DAG* direpresentasikan kedalam kromosom biner. Susunannya digambarkan seperti pada Gambar 5.

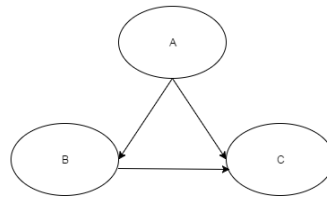


Gambar 5. Pembagian representasi gen pada kromosom.

Kromosom terbagi kedalam 2 bagian utama yaitu bagian kiri yang menentukan urutan variabel pada matriks konektivitas dan bagian kanan yang menentukan apakah sebuah variabel terhubung atau tidak. Penentuan urutan variabel ditentukan berdasarkan hasil pengurutan konversi nilai order bit masing-masing variabel setelah diubah kedalam bentuk bilangan integer basis 10. Apabila order bit bernilai sama, maka urutannya akan ditentukan secara *random*. Pada bit yang mendefinisikan matriks konektivitas, nilai 1 berarti terhubung, dan nilai 0 berarti tidak terhubung. Untuk menjamin tidak adanya siklus pada graf yang dibentuk maka pada matriks konektivitas, *cell* yang hanya dapat diisi adalah *cell* yang terdapat pada setengah matriks bagian bawah yang terbagi oleh diagonal pada pertengahan matriks atau yang membentuk sebuah segitiga bagian bawah matriks. Urutan yang ditentukan oleh order bit yang dimaksud adalah posisi baris untuk masing-masing variabel. Variabel dengan order bit paling tinggi akan diletakkan pada baris pertama. Struktur *BN* hasil translasi dari kromosom pada Gambar 5 digambarkan pada Gambar 7 dengan matriks konektivitasnya dijelaskan pada Gambar 6.

	Parent		
	A	B	C
A	-	-	-
B	1	-	-
C	1	1	-

Gambar 6. Hasil translasi contoh kromosom kedalam matriks konektivitas.



Gambar 7. Hasil translasi contoh kromosom kedalam DAG.

E. Pengisian CPD

Untuk dapat menghitung skor $BDeu$, dibutuhkan nilai-nilai setiap jenis nilai masing-masing variabel dengan diketahui setiap jenis nilai *parentnya*. Pengisian CPD ini dilakukan berdasarkan data set yang dijadikan masukan.

F. Pengisian CPD

Formula untuk evaluasi kromosom disesuaikan dengan formula skor yang digunakan pada penelitian yang telah ada sebelumnya untuk mempermudah *benchmarking*, yaitu skor $BDeu$. Untuk skor jenis ini, terdapat 1 buah parameter perhitungan yaitu nilai ESS yang nantinya akan dilibatkan pada proses penghitungan. Untuk penjelasan terkait proses penghitungannya, dapat dilihat pada bagian Evaluasi *Bayesian Networks*.

G. Terminasi Program

Pada penentuan apakah proses evolusi akan tetap dijalankan atau tidak, terdapat dua pilihan patokan. Yang pertama apabila salah satu kromosom telah mencapai suatu skor $BDeu$ yang ditentukan maka proses evolusi akan berhenti, atau apabila proses evolusi telah mencapai jumlah generasi yang ditentukan maka proses evolusi akan berhenti. Sistem yang dirancang menggunakan terminasi berdasarkan jumlah generasi. Jumlah generasi yang ditentukan menjadi salah satu parameter pengaturan.

IV. PENGUJIAN DAN ANALISIS

A. Tujuan Pengujian

Tujuan dilakukannya pengujian pada sistem adalah sebagai berikut:

- 1) Mendapatkan nilai evaluasi skor $BDeu$ dari solusi berbentuk susunan DAG yang dikeluarkan sistem.
- 2) Melihat efek dari penyesuaian masing-masing parameter pengaturan.
- 3) Melakukan perbandingan nilai evaluasi skor $BDeu$ terhadap hasil pengujian *GOBNILP*.

B. Data yang diujikan

Terdapat 5 buah data set yang diujikan, semua data tersebut diperoleh dari situs web *GOBNILP* <https://www.cs.york.ac.uk/aig/sw/gobnilp/data/> [5]. Namun karena keterbatasan formula *BDeu default*, sistem tidak dapat menerima data set dengan jumlah baris lebih dari 171 baris. Rincian data set yang diujikan beserta skor *BDeu* yang diperoleh untuk masing-masing data set oleh *GOBNILP* dijelaskan pada Tabel 1.

TABEL 1
RINCIAN DATA SET YANG DIJADIKAN OBJEK PENGUJIAN

Data set	Jumlah variabel	Jumlah baris	Skor <i>BDeu GOBNILP</i> [6]
Alarm_100	37	100	-1349.23
Asia_100	8	100	-245.64
Carpo_100	56	100	-1829.31
Water_100	26	100	-1500.97
Insurance_100	26	100	-1686.23

1) *Alarm 100*

ALARM (A Logical Alarm Reduction Mechanism) merupakan sistem alarm pesan untuk memonitor pasien. Sistem mengkalkulasi probabilitas diagnosis berdasarkan gejala-gejala yang ada. Variabel medis yang terlibat meliputi 8 diagnosis, 16 *findings*, dan 13 variabel *intermediate* [6].

2) *Asia 100*

Data set ini menggambarkan hubungan antara apakah seseorang mengunjungi benua asia, terkena *tuberculosis*, terkena kanker paru-paru, merokok, terkena *bronchitis*, terkena *dyspnoe*, dan memiliki hasil positif terkena kanker pada X-ray [7].

3) *Water 100*

Merupakan sistem yang memodelkan manajemen dan perawatan limbah air [9].

4) *Insurance 100*

Merupakan jaringan yang digunakan untuk mengestimasi biaya klaim asuransi mobil [10].

C. *Skenario Pengujian*

Dilakukan 10 macam kombinasi parameter pengujian seperti pada Tabel 2 untuk setiap data set yang diujikan. Keberagaman kombinasi parameter pengujian ini ditujukan untuk mengobservasi kombinasi parameter mana yang menghasilkan skor *BDeu* paling baik. Setiap skenario pengujian dilakukan sebanyak 30 kali dikarenakan terdapat unsur keacakan pada sistem probabilistik yang diuji.

Pada setiap observasi pengaturan parameter, dilakukan analisis dampak dari pengaturan parameter yang sedang diobservasi terhadap skor *BDeu* yang dihasilkan.

Scoring function yang dipakai pada penelitian ini untuk menilai apakah *BN* yang dihasilkan adalah baik atau tidak adalah skor *BDeu*. Dari 30 hasil pengujian pada setiap kombinasi parameter pada setiap data set, skor *BDeu* yang diperoleh diambil nilai maksimum, nilai rata-rata, dan nilai minimumnya. Nilai maksimum dari seluruh hasil pengujian pada setiap data set akan diambil yang berikutnya akan dibandingkan dengan skor *BDeu* yang dihasilkan oleh *GOBNILP*.

TABEL 2
KOMBINASI PARAMETER PENGUJIAN YANG DILAKUKAN

POPs	GENs	f	b	cr	ob	ESS
25	800	2	10	0.9	10	1
25	800	2	5	0.1	10	1
25	800	1	5	0.5	10	1
25	800	1	0	0.5	10	1
25	800	0.5	5	0.1	10	1
100	200	2	10	0.9	10	1
100	200	2	5	0.1	10	1
100	200	1	5	0.5	10	1
100	200	1	0	0.5	10	1
100	200	0.5	5	0.1	10	1

Keterangan

1. POPs : jumlah kromosom pada populasi, parameter ini berfungsi untuk mengatur seberapa ketat persaingan antar kromosom untuk generasi berikutnya dan mengatur kadar keberagaman kromosom pada populasi.
2. GENs : jumlah generasi yang ingin dicapai pada proses evolusi.
3. f : parameter scaling factor NMBDE.
4. b : parameter bandwidth factor NMBDE.
5. cr : probabilitas dilakukannya mutasi lalu rekombinasi pada NMBDE.
6. ob : jumlah order bit untuk masing-masing variabel untuk penentuan urutan variabel pada matriks konektivitas.
7. ESS : parameter scaling skor BDeu

D. Hasil Pengujian

Hasil pengujian berupa skor $BDeu$ untuk setiap data set berdasarkan skenario pengujian yang telah ditetapkan dijadikan landasan analisis terhadap hasil yang diperoleh.

TABEL 3
HASIL PENGUJIAN DATA SET ALARM_100

POPs	GENs	f	b	cr	ob	ess	Max BDeu	Ave BDeu	Min BDeu
25	800	2	10	0.9	10	1	-2124.30	-2199.94	-2263.39
25	800	2	5	0.1	10	1	-1992.65	-2053.61	-2111.68
25	800	1	5	0.5	10	1	-2113.43	-2156.31	-2191.94
25	800	1	0	0.5	10	1	-2076.01	-2160.11	-2215.44
25	800	0.5	5	0.1	10	1	-1973.77	-2059.98	-2093.01
100	200	2	10	0.9	10	1	-2110.25	-2179.05	-2221.31
100	200	2	5	0.1	10	1	-1984.20	-2047.53	-2085.96
100	200	1	5	0.5	10	1	-2081.24	-2135.00	-2179.39
100	200	1	0	0.5	10	1	-2079.33	-2133.72	-2181.69
100	200	0.5	5	0.1	10	1	-2010.00	-2055.79	-2086.19

Setelah dilakukan observasi pengaturan parameter pada pengujian data set Alarm_100 yang dicatat pada Tabel 3, skor $BDeu$ terbaik yang diperoleh bernilai -1973.77 yang didapat dari kombinasi parameter ke 5, yaitu kombinasi yang memiliki parameter f terkecil 0.5, dan cr terkecil 0.1. Dari observasi ini muncul hipotesis bahwa semakin kecil nilai f dan cr , maka semakin baik skor $BDeu$ yang diperoleh. Hipotesis ini akan dicoba diperkuat pada pengujian-pengujian data set lain berikutnya.

TABEL 4
 HASIL PENGUJIAN DATA SET ASIA_100

POPs	GENs	f	b	cr	ob	ess	Max BDeu	Ave BDeu	Min BDeu
25	800	2	10	0.9	10	1	-248.89	-253.09	-261.54
25	800	2	5	0.1	10	1	-243.68	-245.85	-247.58
25	800	1	5	0.5	10	1	-244.63	-247.77	-250.3
25	800	1	0	0.5	10	1	-245.31	-248.12	-253.25
25	800	0.5	5	0.1	10	1	-243.68	-246.37	-248.89
100	200	2	10	0.9	10	1	-245.53	-251.37	-257.29
100	200	2	5	0.1	10	1	-243.68	-245.24	-247.01
100	200	1	5	0.5	10	1	-243.68	-246.6	-248.81
100	200	1	0	0.5	10	1	-243.68	-246.92	-250.38
100	200	0.5	5	0.1	10	1	-243.68	-244.88	-247.23

Pada pengujian data set Asia_100 yang dicatat pada Tabel 4, skor maksimum yang diperoleh adalah -243.68 yang dihasilkan oleh beberapa kombinasi parameter. Hipotesis sebelumnya yang menduga semakin kecil nilai f maka semakin baik skor *BDeu* tidak terbukti dikarenakan ada kombinasi parameter yang memiliki nilai f paling besar yaitu 2 justru memberikan hasil skor *BDeu* yang terbaik diantara observasi yang dilakukan. Namun hipotesis yang menduga semakin kecil nilai cr maka semakin baik skor *BDeu*, ternyata terbukti. Hal ini dikarenakan pada setiap kombinasi parameter yang memiliki nilai cr terkecil yaitu 0.1 menghasilkan skor *BDeu* terbaik.

TABEL 5
 HASIL PENGUJIAN DATA SET CARPO_100

POPs	GENs	f	b	cr	ob	ess	Max BDeu	Ave BDeu	Min BDeu
25	800	2	10	0.9	10	1	-2549.54	-2592.41	-2624.30
25	800	2	5	0.1	10	1	-2450.54	-2495.96	-2527.04
25	800	1	5	0.5	10	1	-2518.41	-2569.86	-2611.36
25	800	1	0	0.5	10	1	-2542.06	-2573.99	-2600.47
25	800	0.5	5	0.1	10	1	-2484.23	-2500.29	-2523.48
100	200	2	10	0.9	10	1	-2511.84	-2569.66	-2592.26
100	200	2	5	0.1	10	1	-2460.23	-2495.02	-2527.39
100	200	1	5	0.5	10	1	-2529.77	-2553.76	-2575.87
100	200	1	0	0.5	10	1	-2523.55	-2555.34	-2582.23
100	200	0.5	5	0.1	10	1	-2458.27	-2497.27	-2527.75

Pada pengujian data set Carpo_100 yang dicatat pada Tabel 5, hipotesis semakin kecil nilai cr maka semakin baik skor *BDeu* telah terbukti. Hal ini terlihat pada Tabel 14 bahwa setiap kali nilai cr diturunkan dari 0.9, ke 0.5, ke 0.1, skor *BDeu* selalu mengalami kenaikan. Serta skor *BDeu* terbaik yang dihasilkan merupakan dari kombinasi parameter yang memiliki nilai cr terkecil yaitu 0.1.

TABEL 6
HASIL PENGUJIAN DATA SET INSURANCE_100

POPs	GENs	f	b	cr	ob	ess	Max BDeu	Ave BDeu	Min BDeu
25	800	2	10	0.9	10	1	-2203.34	-2248.15	-2296.94
25	800	2	5	0.1	10	1	-2061.05	-2113.48	-2157.68
25	800	1	5	0.5	10	1	-2141.09	-2205.11	-2247.74
25	800	1	0	0.5	10	1	-2157.36	-2208.40	-2259.74
25	800	0.5	5	0.1	10	1	-2024.17	-2114.39	-2147.16
100	200	2	10	0.9	10	1	-2157.92	-2206.74	-2249.91
100	200	2	5	0.1	10	1	-2047.32	-2102.96	-2138.98
100	200	1	5	0.5	10	1	-2145.49	-2189.25	-2221.71
100	200	1	0	0.5	10	1	-2146.13	-2185.14	-2220.86
100	200	0.5	5	0.1	10	1	-2055.55	-2104.59	-2136.71

Pada pengujian data set Insurance_100 yang dicatat pada Tabel 6, hipotesis nilai cr berbanding terbalik dengan skor BDeu telah terbukti. Hal ini terlihat pada setiap penurunan nilai cr terbukti skor BDeu semakin naik.

TABEL 7
HASIL PENGUJIAN DATA SET WATER_100

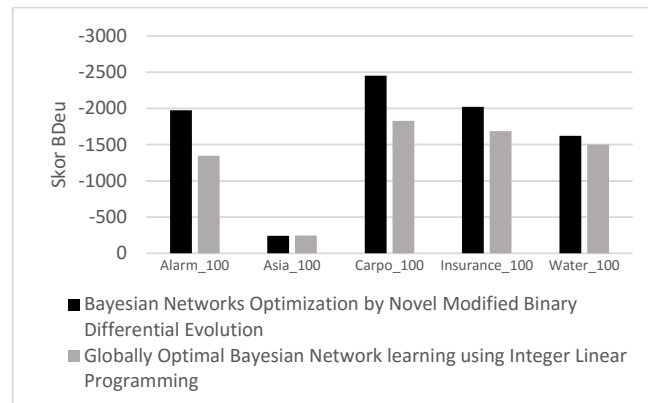
POPs	GENs	f	b	cr	ob	ess	Max BDeu	Ave BDeu	Min BDeu
25	800	2	10	0.9	10	1	-1728.05	-1760.50	-1794.89
25	800	2	5	0.1	10	1	-1639.27	-1674.20	-1706.64
25	800	1	5	0.5	10	1	-1678.62	-1738.79	-1776.27
25	800	1	0	0.5	10	1	-1699.00	-1739.44	-1771.57
25	800	0.5	5	0.1	10	1	-1627.82	-1680.14	-1714.26
100	200	2	10	0.9	10	1	-1702.19	-1746.68	-1770.50
100	200	2	5	0.1	10	1	-1641.05	-1671.61	-1696.04
100	200	1	5	0.5	10	1	-1672.62	-1720.27	-1754.86
100	200	1	0	0.5	10	1	-1690.98	-1723.91	-1745.09
100	200	0.5	5	0.1	10	1	-1621.90	-1665.53	-1689.31

Pada pengujian data set Water_100 yang dicatat pada Tabel 7, hipotesis nilai cr berbanding terbalik dengan skor BDeu yang dihasilkan dapat dikatakan terbukti. Pada setiap penurunan nilai cr, selalu disertai kenaikan skor BDeu, dan skor BDeu maksimum yang diperoleh pada pengujian data set ini juga didapat dari kombinasi parameter yang memiliki nilai cr paling kecil yaitu 0.1. Hal ini sama seperti pada data set – data set sebelumnya.

E. Hasil Pengujian

TABEL 8
PERBANDINGAN SKOR BDEU YANG DIPEROLEH DENGAN SKOR BDEU GOBNILP

	Alarm_100	Asia_100	Carmo_100	Insurance_100	Water_100
Bayesian Networks Optimization by Novel Modified Binary Differential Evolution	-1973.77	-243.68	-2450.54	-2024.17	-1621.90
Globally Optimal Bayesian Network learning using Integer Linear Programming [6]	-1349.23	-245.64	-1829.31	-1686.23	-1500.97



Gambar 8. Grafik perbandingan skor *BDeu* yang diperoleh dengan skor *BDeu* GOBNILP.

Setelah didapatkan hasil dari pengujian, hasil maksimum dari setiap data set yang diujikan dibandingkan dengan hasil dari GOBNILP dengan data set yang sama. Tabel 8 memperlihatkan perbandingan hasil skor *BDeu* yang diperoleh, dan Gambar 8 mengilustrasikan perbandingan tersebut kedalam diagram batang.

Skor *BDeu* yang diukur adalah fungsi maksimasi, yang artinya semakin tinggi nilai maka semakin baik. Semakin baik disini berarti *BN* yang dihasilkan semakin baik dalam memodelkan data yang diberikan.

F. Analisis Hasil Pengujian

Dari keseluruhan data set yang diujikan, kombinasi parameter ke-5 dengan POPs = 25, GENs = 800, f = 0.5, b = 5, cr = 0.1, ob = 10, ESS = 1 dapat menghasilkan hasil terbaik untuk 3 buah data set. Walaupun setiap data set memiliki kombinasi parameter terbaiknya masing-masing, namun ini dapat dijadikan acuan untuk pengujian berikutnya dengan memulai observasi kombinasi parameter dimulai dari titik awalnya adalah kombinasi parameter ke-5 tadi.

Pada setiap observasi pengaturan setiap parameter, hanya parameter *cr* yang memberikan dampak signifikan pada setiap perubahannya. Parameter yang lain hanya memberikan dampak *minor* yang menyebabkan tidak dapat ditarik kesimpulan dari dampak kenaikan atau penurunan nilainya. Pada setiap penurunan nilai parameter *cr*, dapat dilihat bahwa skor *BDeu* yang dihasilkan semakin baik. Hal ini menunjukkan semakin jarang dilakukan mutasi dan rekombinasi pada proses evolusi maka skor *BDeu* yang dihasilkan semakin baik.

Performansi sistem yang diuji lebih rendah dari performansi GOBNILP. Hal ini dikarenakan sistem yang diuji bersifat probabilistik yang pada setiap pengujian solusi yang didapatkan tidak pasti. Semakin besar ruang solusi maka semakin besar pula jumlah evaluasi kromosom yang diperlukan, dan semakin banyak pula kemungkinan solusi yang dapat dihasilkan. Sedangkan GOBNILP bersifat deterministik yang dapat menemukan solusi pasti untuk setiap pengujiannya. Hal ini terlihat dampaknya pada data set yang memiliki jumlah variabel sedikit, yaitu Asia_100 yang hanya memiliki 8 buah variabel terlihat bahwa sistem yang diuji unggul dari GOBNILP dikarenakan sistem dapat mengevaluasi sebagian banyak solusi yang ada pada ruang solusi.

V. KESIMPULAN DAN SARAN

Probabilitas mutasi dan rekombinasi memiliki dampak tertinggi pada solusi yang dihasilkan. Semakin rendah probabilitas mutasi dan rekombinasi maka semakin baik skor *BDeu* dari solusi yang dihasilkan. Parameter-parameter yang lain hanya memberikan dampak *minor* sehingga tidak dapat ditaksir dampaknya terhadap skor *BDeu* dari solusi yang dihasilkan.

Sistem yang dibuat dapat lebih unggul dari algoritma GOBNILP pada pengujian data set yang memiliki jumlah variabel kecil. Skor *BDeu* untuk solusi yang dihasilkan untuk data set Asia_100 bernilai -243.68392 lebih tinggi sebesar 1.960344 dari skor *BDeu* dari solusi yang dihasilkan GOBNILP yang bernilai -245.644264.

Performansi sistem probabilistik yang dibuat dapat lebih baik apabila jumlah evaluasi kromosom yang dilakukan sebanding dengan jumlah ruang solusi yang ingin dilakukan pencarian. Pada data set Asia_100 yang memiliki ruang solusi kecil, terbukti sistem dapat menghasilkan solusi yang lebih baik dari *GOBNILP*.

Sistem probabilistik seperti sistem yang dibuat ini cocok untuk menangani masalah yang membutuhkan aproksimasi ke solusi optimal, namun memiliki batasan memori dan waktu. Karena sejauh mana aproksimasi yang dilakukan dapat diatur dengan parameter jumlah evaluasi kromosom yaitu parameter jumlah kromosom pada populasi dan jumlah generasi yang ingin ditempuh.

Besar ruang solusi sistem yang diajukan lebih besar dari *GOBNILP*. Hal ini berarti dengan perangkat dan waktu yang memadai maka hasil keluaran sistem secara teori pasti dapat lebih baik dari *GOBNILP*, hal ini dapat dicapai apabila parameter jumlah evaluasi kromosom yang ditetapkan menggunakan nilai pada angka yang sangat besar atau terminasi program dilakukan apabila skor *BDeu* yang diperoleh telah melampaui skor *BDeu* milik *GOBNILP*.

DAFTAR PUSTAKA

- [1] D. Koller and N. Friedman, *Probabilistic Graphical Models Principles and Techniques*, The MIT Press.
- [2] H. Xing-Chen, Q. Zheng and S. L.-P. Tian Lei, "Learning Bayesian Network Structures with Discrete Particle Swarm Optimization," *IEEE Symposium on Foundations of Computational Intelligence (FOCI 2007)*, 2007.
- [3] L. Wang, X. Fu, Y. Mao, M. I. Menhas and M. Fei, "A novel modified binary differential evolution algorithm and its applications," *Bio-inspired computing and applications (LSMS-ICSEE '2010)*, vol. 98, pp. 55-75, 2012.
- [4] M. Bartlett and J. Cussens, "Advances in Bayesian Network Learning using Integer Programming," *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI 2013)*, 2013.
- [5] B. Malone, "Scoring Functions for Learning Bayesian Networks," Lecture notes of Probabilistic Graphical Models at Department of Computer Science University of Helsinki, 2014.
- [6] J. Cussens and M. Bartlett, "Globally Optimal Bayesian Network learning using Integer Linear Programming," [Online]. Available: <https://www.cs.york.ac.uk/aig/sw/gobnilp/>.
- [7] H. J. Suermondt, R. M. Chavez, G. F. Cooper, M.D., Ph.D. and I. A. Belnlich, M.D., "The ALARM Monitoring System: A Case Studi with two Probabilistic Inference Techniques for Belief Networks," *Medical Informatics*.
- [8] S. Lauritzen and D. Spiegelhalter, "Local Computations with Probabilities on Graphical Structures and Their Application to Expert Sustems," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 50, no. 2, pp. 157-224, 1988.
- [9] F.V. Jensen, U. Kjærulff, K.G. Olesen and J. Pedersen, "Et Forprojekt Til et Ekspertsystem for Drift af Spildevandsrensning (An Expert System for Control of Waste Water Treatment - A Pilot Project)," *Technical Report, Judex Datasystemer A/S, Aalborg*, 1989.
- [10] J. Binder, D. Koller, S. Russel and K. Kanazawa, "Adaptive Probabilistic Networks with Hidden Variables," *Machine Learning*, pp. 213-244, 1998.

