

Prediksi Penyebaran Penyakit Demam Berdarah Dengue (DBD) di Kabupaten Bandung menggunakan Hybrid *Random Forest* (RF) dan *Genetic Algorithm* (GA)

Maharani Padma Utami¹, Fhira Nhita, M.T², Annisa Aditsania, M.Si³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹padmamaharany@students.telkomuniversity.ac.id, ²fhiranhita@telkomuniversity.ac.id,

³aaditsania@telkomuniversity.ac.id

Abstrak

Penyakit Demam Berdarah Dengue (DBD) merupakan penyakit yang ditularkan melalui nyamuk *Aedes Aegypti*. Nyamuk tersebut berada di wilayah tropis dan sub tropis seperti Asia Tenggara. Di Indonesia, khususnya Kabupaten Bandung jumlah kasus penyakit ini cenderung meningkat dan daerah penyebarannya semakin luas. Oleh karena itu, diperlukan informasi mengenai penyebaran untuk membantu menganalisa dalam menentukan tindakan yang harus dilakukan untuk menanganinya pada masa yang akan datang. Dalam menyusun informasi ini, dilakukan pengolahan data penyinaran lama matahari, kelembapan, angin, *temperature*, uap, curah hujan, jumlah kasus, jumlah penduduk dan presentase PHBS di Kabupaten Bandung menggunakan metode *Hybrid Random Forest* dan *Genetic Algorithm* untuk menemukan prediksi penyebaran penyakit yang hasilnya ditampilkan dengan *tools* yaitu GIS (*Geographic Information System*). Perbandingan menggunakan algoritma *Random Forest* tanpa *Genetic Algorithm* dan *Random Forest* dengan *Genetic Algorithm*, adalah akurasi rata-rata sebesar 83,42% dan 89,12 %.

Kata kunci : prediksi, penyebaran, demam berdarah *dengue*, *Random Forest*, *Genetic Algorithm*

Abstract

Dengue Hemorrhagic Fever (DHF) is a disease transmitted through the *Aedes Aegypti* mosquito. These mosquitoes are in tropical and sub-tropical regions such as Southeast Asia. In Indonesia, especially Bandung Regency, the number of cases of disease has increased and the area of spread is increasingly widespread. Therefore, information about dissemination is needed that helps analyze in determining the actions that must be taken to handle it in the future. In compiling this information, data processing of solar radiation, humidity, wind, temperature, wind, rainfall, number of cases, population and percentage of PHBS in Bandung Regency used *Random Forest Hybrid* methods and *Genetic Algorithms* to find out how much was needed by findings with tools, namely GIS (*Geographic Information System*). Comparison using *Random Forest* algorithm without *Genetic Algorithm* and *Random Forest* with *Genetic Algorithm*, is an average accuracy of 83.42% and 89.12%.

Keywords: prediction, spreading, dengue hemorrhagic fever, *Random Forest*, *Genetic Algorithm*

1. Pendahuluan

Pada bab ini, diberikan penjelasan dari abstrak dengan beberapa tambahan. Isi utama dari bab ini ialah latar belakang, penjelasan masalah, dan tujuan.

1.1 Latar Belakang

Demam Berdarah *Dengue* (DBD) hingga saat ini termasuk dalam salah satu permasalahan kesehatan masyarakat di Indonesia karena seluruh provinsi telah terjangkit oleh penyakit ini. Sebagai negara dengan iklim tropis yang paling besar di Asia Tenggara. Pada daerah dengan iklim tersebut merupakan tempat berkembangnya nyamuk *Aedes Aegypti*. Bersamaan dengan kondisi Negara Indonesia yang mengalami kepadatan penduduk, bertambahnya jumlah penderita DBD, perubahan iklim dan perluasan daerah. Pada tahun 2016, tercatat sebanyak 204.171 penderita DBD di 34 provinsi di Indonesia dan 1.598 orang diantaranya meninggal dunia[1]. Jumlah tersebut lebih tinggi dibanding tahun sebelumnya, yaitu sebanyak 129.650 penderita DBD dan sebanyak 1.071 penderita meninggal dunia pada tahun 2015[2]. Hal ini mengartikan terjadi peningkatan pada tahun 2015 menuju 2016 sebanyak 57,47% pada jumlah penderita dan 49,20% pada jumlah meninggal.

Indonesia khususnya Kabupaten Bandung yang terdiri dari 31 Kecamatan mengalami fluktuasi dari tahun 2014 sampai dengan tahun 2018, tercatat sebanyak 995 penderita dengan 4 orang penderita meninggal dunia pada tahun 2014, sebanyak 1.013 penderita dengan 4 orang penderita meninggal dunia pada tahun 2015, sebanyak 3.470 penderita dengan 10 orang penderita meninggal dunia pada tahun 2016, 1.015 penderita dengan 4 orang penderita meninggal dunia pada tahun 2017, serta sebanyak 1.778 penderita dengan 11 orang penderita meninggal dunia pada tahun 2018[3]. Dari data tersebut dapat diketahui bahwa peningkatan paling tinggi yaitu pada tahun 2016.

Dengan demikian diperlukan penanganan segera terhadap penyebaran penyakit Demam Berdarah di kabupaten Bandung. Salah satu yang dapat dilakukan ialah memprediksi penyebarannya dimasa mendatang sebagai bahan evaluasi dari upaya-upaya pencegahan yang telah dilakukan seperti Pola Hidup Bersih Sehat (PHBS) maupun persiapan dalam menangani penyakit tersebut. Adapun dalam melakukan prediksi tersebut data kabupaten Bandung dikelompokkan menjadi 7 daerah yang terdiri dari beberapa kecamatan yaitu Regional 1 meliputi Kecamatan Ciwidey, Kutawaringin, Pasirjambu, Rancabali, dan Soreang. Regional 2 meliputi Kecamatan Dayeuhkolot, Katapang, Margaasih, dan Margahayu. Regional 3 meliputi Kecamatan Bojongsoang, Cilengkrang, Cileunyi, dan Cimencyan. Regional 4 meliputi Kecamatan Cicalengka, Cikancung, Nagreg, dan Rancaekek. Regional 5 meliputi Kecamatan Ibum, Majalaya, Paseh, dan Solokanjeruk. Regional 6 meliputi Kecamatan Baleendah, Ciparay, Kertasari, dan Pacet. Serta Regional 7 meliputi Kecamatan Arjasari, Banjaran, Cangkuang, Cimaung, Pameungpeuk dan Pengalengan[4].

Pada penelitian sebelumnya, dilakukan klasifikasi riset ahli dan pemula berdasarkan faktor gaya kognitif dalam pengaturan akademik menggunakan beberapa teknik *decision tree* dan salah satunya *Random Forest* yang memberi akurasi terbaik yaitu 92,73 yang menggunakan data mahasiswa penelitian UTM sebanyak 34 mahasiswa penelitian master dan 76 mahasiswa PHD dari berbagai fakultas[5]. Dalam melakukan metode klasifikasi seperti yang telah dilakukan pada penelitian tersebut. Penulis menemukan bahwa metode *Random Forest* memiliki akurasi yang lebih tinggi.

Adapun faktor-faktor yang disebutkan berpengaruh pada beberapa penelitian tentang Demam Berdarah *Dengue* sebelumnya yaitu cuaca, suhu dan kepadatan penduduk dapat[6]. Disebutkan dalam salah satu Jurnal kesehatan curah hujan, suhu dan kelembaban ialah iklim yang paling berpengaruh[7]. Berdasarkan faktor-faktor yang mempengaruhi tersebut dilakukan penelitian menggunakan 12 atribut yaitu penyinaran lama matahari, kelembapan, angin, suhu, uap, curah hujan, jumlah kasus, meninggal, penduduk, PHBS, IR dan CFR. Dalam hal ini, perlu diketahui faktor mana saja yang paling berpengaruh sehingga dilakukan proses seleksi menggunakan *Genetic Algorithm* (GA).

1.2 Perumusan Masalah

Perumusan masalah pada penelitian ini ialah menganalisis hubungan antara atribut data cuaca dan kasus penyakit DBD terhadap penyebaran penyakit DBD di Bandung, menganalisis pengaruh seleksi atribut dalam memprediksi penyebaran penyakit DBD di Kabupaten Bandung. Adapun batasan masalah berupa data yang digunakan ialah lama penyinaran matahari, kelembapan, angin, data suhu rata-rata, uap, dan curah hujan dari BMKG (Badan Meteorologi, Klimatologi dan Geofisika) Kabupaten Bandung mulai dari 2014 hingga 2017, data jumlah penduduk dari BPS (Badan Pusat Statistika) Kabupaten Bandung. dan data jumlah kasus DBD dan PHBS dari Dinas Kesehatan Kabupaten Bandung.

1.3 Tujuan

Tujuan yang ingin dicapai dari penelitian ini adalah dapat melakukan prediksi, menganalisa perbandingan dan hubungan atribut data terhadap penyebaran penyakit DBD di Kabupaten Bandung antara menggunakan Algoritma *Random Forest* (RF) tanpa *Genetic Algorithm* (GA) dan Hybrid Algoritma *Random Forest* (RF) dan *Genetic Algorithm* (GA).

2. Studi Terkait

2.1 Demam Berdarah Dengue (DBD)

Terdapat beberapa penelitian yang telah dilakukan terkait penyakit ini. Pada tahun 2015, dilakukan penelitian untuk mendapatkan gambaran pola sebaran kasus DBD dengan penggunaan lahan melalui pendekatan Sistem Informasi Geografi berdasarkan karakteristik penderita, faktor lingkungan dan pola penggunaan lahan pemukiman di Kabupaten Banggai, Sulawesi Tengah. Penelitian ini menyimpulkan bahwa terdapat keterkaitan antara kepadatan penduduk dan ketinggian wilayah terhadap kasus DBD serta rata-rata curah hujan, kelembaban dan suhu mendukung tempat perkembangbiakan umur nyamuk *Aedes Aegypti*[8]. Disebutkan pula pada penelitian lain bahwa terdapat keterkaitan antara cuaca dan penyakit DBD sehingga dapat dilakukan prediksi penyebarannya[9]. Serta pada tahun 2018, dilakukan penelitian untuk memperoleh pusat titik penyebaran penyakit demam berdarah di kota Tangerang selatan dengan melakukan pengelompokan penyebaran menggunakan algoritma K-Means[10].

2.2 Random Forest

Random Forest merupakan salah satu teknik klasifikasi data dalam jumlah yang besar dan dikenal memiliki kinerja sangat baik jika dibandingkan dengan teknik klasifikasi lainnya, karena sangat mudah digunakan dan memberikan performansi akurasi yang lebih tinggi[5]. Metode ini juga bisa digunakan untuk regresi maupun tujuan lainnya. *Random Forest* pertama kali dipublikasikan melalui makalah oleh Leo Breiman pada tahun 2001[11]. Adapun beberapa penelitian yang telah dilakukan dengan RF dirangkum dalam tabel 1.

Tabel 1. Perbandingan Akurasi Penelitian menggunakan Algoritma *Random Forest*

Referensi	Algoritma	Data	Akurasi
[12]	<i>Random Forest (RF)</i>	Data sampel darah demam berdarah dengue, yang dikumpulkan dari Rumah sakit Holy Family, Rawalpindi, Pakistan.	91%
[8]	LMT	Data mahasiswa penelitian UTM yang terdiri dari 34 mahasiswa penelitian master dan 76 mahasiswa PHD dari berbagai fakultas	91.81%
	J48		92.72%
	REP tree		90.00%
	<i>Random Forest (RF)</i>		92.73%
	<i>Decision Stump</i>		74.55%
	<i>Random Tree</i>		90.91%
[13]	<i>Back propagation neural network (BPNN)</i>	Data sensor respon dari lidah sebagai fitur, dengan jus jeruk dan cuka cina sebagai sampel	86.68%
	<i>Support Vector Machine (SVM)</i>		66.45%
	<i>Random Forest (RF)</i>		99.07%

Dari perbandingan data akurasi pada beberapa penelitian diatas dapat disimpulkan bahwa dalam kasus-kasus tersebut *Random Forest (RF)* memiliki keunggulan dibanding metode atau algoritma lainnya.

2.3 Genetic Algorithm

Pertama kali *Genetic Algorithm (GA)* diusulkan oleh John Holland pada tahun 1975 pada buku yang berjudul *Adaptation in Natural and Artificial Systems*[14]. Algoritma ini memanfaatkan proses seleksi alamiah

yang dikenal dengan proses evolusi yaitu suatu individu selalu mengalami perubahan gen untuk menyesuaikan dengan lingkungannya. Dalam implementasinya algoritma ini telah banyak diaplikasikan melalui berbagai bidang. Salah satunya optimisasi rute perjalanan dengan meminimisasi jarak tempuh untuk keseluruhan rute perjalanan antar kota [15]. Dilakukan pula perencanaan jalur robot pembersih, dengan posisi robot sebagai gen dan jalur-jalur kecil sebagai kromosom [16]. Penggabungan GA dengan algoritma lainnya juga telah dilakukan, yaitu penggabungan GA dengan algoritma *backpropagation* untuk menghitung nilai MSE dalam perhitungan fitness pada GA untuk pengujian menemukan prediksi waktu panen tebu [17]. Sedangkan GA yang digunakan untuk seleksi fitur telah dilakukan dalam beberapa penelitian pada tabel berikut.

Tabel 2. Perbandingan Akurasi Penelitian menggunakan Algoritma Genetika

Referensi	Algoritma	Akurasi
[18]	<i>SVM</i>	90%
	<i>SVM & GA</i>	87,857%
[19]	Naïve Bayes	60%
	Naïve Bayes & GA	89,50%
[20]	Naïve Bayes	63,69%
	Naïve Bayes & GA	82,67%
[21]	C4.5	83,25%
	C4.5 & GA	86,47%

Dari perbandingan data akurasi pada beberapa penelitian diatas sebagian besar algoritma yang digabungkan dengan GA memiliki akurasi yang lebih tinggi dibanding tanpa menggunakan GA.

2.4 Confusion Matrix

Melakukan pengukuran kinerja dari suatu sistem klasifikasi merupakan hal yang penting untuk menggambarkan seberapa baik sistem dalam mengklasifikasikan data. Salah satu metodenya yaitu *Confusion Matrix*. Pada metode ini mengandung informasi membandingkan hasil klasifikasi yang seharusnya. Terdapat empat istilah sebagai representasi hasil proses klasifikasi yaitu

- *True Positive* (TP) yaitu jumlah data positif terdeteksi benar,
- *True Negative* (TN) yaitu jumlah data negatif terdeteksi benar,
- *False Positive* (FP) yaitu jumlah data positif terdeteksi salah,
- *False Negative* (FN) yaitu jumlah data negatif terdeteksi salah,

Confusion matrix dapat digambarkan pada tabel seperti dibawah ini

Tabel 3. Confusion matrix

		Nilai Sebenarnya	
		<i>True</i>	<i>False</i>
Nilai prediksi	<i>True</i>	TP (<i>True Positive</i>) <i>Correct result</i>	FP (<i>False Positive</i>) <i>Unexpected result</i>
	<i>False</i>	FN (<i>False Negative</i>) <i>Missing result</i>	TN (<i>True Negative</i>) <i>Correct absence of result</i>

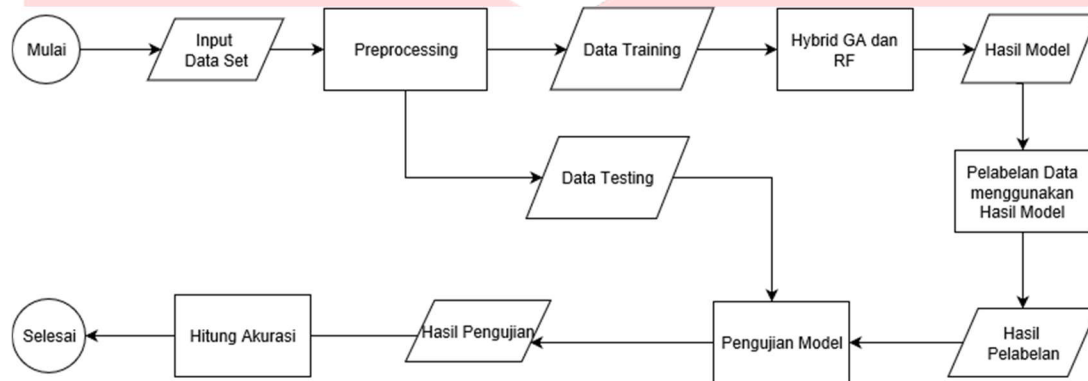
Pada kasus ini, data negatif yang dimaksud ialah data yang bernilai 1 pada kelas. Sedangkan data positif ialah data yang bernilai 2 pada kelas.

Berdasarkan nilai-nilai tersebut dapat diperoleh nilai akurasi yang menggambarkan nilai keakuratan sistem untuk dapat mengklasifikasikan data secara benar atau perbandingan antara data terklasifikasi benar dengan seluruh data yang dirumuskan:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

3. Sistem yang Dibangun

Pada bagian ini memaparkan mengenai rancangan sistem yang telah dibuat atau digunakan pada penelitian.



Gambar 1. Skema Sistem

3.1 Data Set

Keseluruhan data yang digunakan ialah data kabupaten Bandung yang terdiri dari 7 Regional selama 47 bulan, artinya setiap data per Regional memiliki 47 record. Data ini terdiri dari 6 atribut cuaca yaitu penyinaran lama matahari, kelembapan, angin, temperaturee, uap dan curah hujan yang didapatkan dari BMKG, 4 atribut penyakit yaitu jumlah kasus, meninggal, penduduk dan PHBS yang didapatkan dari Dinas Kesehatan Kabupaten Bandung serta 1 atribut turunan yaitu IR.

Tabel 4. Data Set

NO	Bulan	Tahun	Penyinaran lama matahari (%)	Kelembapan (%)	Angin (knot)	Temperature (C)	Uap (mm)	Curah Hujan (mm)	Jumlah Kasus	Penduduk	PHBS (%)	IR
1	JANUARI	2014	42	82	4	22.5	3.3	309	10	443,460	45.18	2,36
2	FEB	2014	47	80	4	22.9	3.6	89	3	443,460	45.18	0,68
3	MARET	2014	52	82	3	23.3	3.4	418.7	14	443,460	45.18	3,20
4	APRIL	2014	58	81	3	23.7	3.7	217.6	7	443,460	45.18	1,67
:	:	:	:	:	:	:	:	:	:	:	:	:
47	November	2017	35	82	5	23.1	3	442	23	619902	50.24	1.08

Atribut IR ialah *incident rate* yang didapat dari perhitungan sebagai berikut :

$$IR = \left(\frac{\text{jumlah kasus}}{\text{penduduk}} \right) \times 100.000 \tag{2}$$

IR adalah rasio kejadian DBD dengan jumlah penduduk per 100.000

3.2 Preprocessing

3.2.1 Pelabelan Kelas

Dilakukan dengan metode Equal-Width :

$$w = \frac{\text{max} - \text{min}}{k} \tag{3}$$

dengan,

w : lebar interval atau nilai pembatas untuk menentukan kelas

max : nilai maksimum

min : nilai minimum

k : jumlah kelas

Hasil pelabelan pada setiap data set sesuai dengan nilai IRnya masing-masing ialah.

Tabel 5. Pelabelan kelas pada setiap data set

Kelas	Reg. 1	Reg. 2	Reg. 3	Reg. 4	Reg. 5	Reg. 6	Reg. 7
1	<=2.88	<=5.49	<=5.16	<=10.65	<=4.79	<=13.14	<=9.36
2	>2.88	>5.49	>5.16	>10.65	>4.79	>13.14	>9.36

3.2.2 Normalisasi

Normalisasi adalah metode untuk mengubah data menjadi skala tertentu. Pada penelitian ini, dilakukan perubahan data menjadi range [0,1] dengan rumus :

$$v' = \frac{v - \min_A}{\max_A - \min_A} \tag{4}$$

dengan,

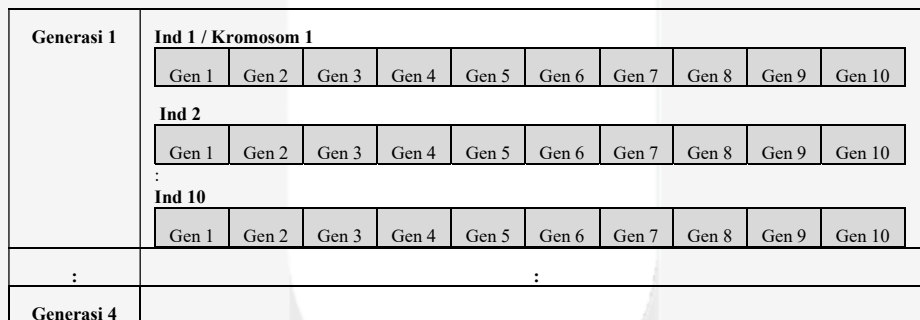
- v' : nilai baru
- v : nilai lama
- \min_A : nilai minimal
- \max_A : nilai maksimal

Tabel 6. Data Set hasil Normalisasi

NO	Penyinaran lama matahari (%)	Kelembapan (%)	Angin (knot)	Temperature (C)	Uap (mm)	Curah Hujan (mm)	Jumlah Kasus	Penduduk	PHBS (%)	IR
1	0.135	0.913	0.33	0.24	0.286	0.466	0.114	0	0.142	0.118
2	0.231	0.826	0.33	0.4	0.4286	0.133	0.033	0	0.142	0.035
3	0.327	0.913	0	0.56	0.333	0.632	0.154	0	0.142	0.160
4	0.442	0.870	0	0.72	0.476	0.328	0.080	0	0.142	0.083
:	:	:	:	:	:	:	:	:	:	:
47	0.404	0.826	0.33	0.6	0.381	0.520	0.155	1	0.820	0.152

3.3 Klasifikasi

3.3.1 Algoritma Genetika : Pendefinisian Individu



Gambar 2. Ilustrasi Populasi pada Algoritma Genetika yang belum berisi nilai

Pada gambar 1 menunjukkan populasi yang terdiri dari 10 individu, setiap individu memiliki 10 *genotype*. Gen (*Genotype*) adalah variable pembentuk kromosom, *Allele* sebagai nilai dari gen tersebut, kromosom merupakan kumpulan dari gen, individu merupakan kumpulan dari kromosom, populasi merupakan sekumpulan individu dan generasi menyatakan satu satuan siklus proses evolusi. Data yang ada dikonversikan dalam bentuk individu yang diwakili oleh satu atau lebih kromosom yang berbentuk biner, yaitu tiap gennya bernilai *True* atau *False*. Pada penelitian ini terdapat 10 fitur sehingga setiap individu terdiri dari 10 kromosom. Dan setiap individu merupakan solusi dari permasalahan dalam penelitian ini. Ukuran Individu berjumlah sama untuk setiap populasi sehingga disebut *static Genetic Algorithm*.

3.3.2 Inisialisasi Kromosom

Pada tahap awal ini, dilakukan inisialisasi kromosom dengan *merandom* angka pada setiap gen/fitur yang akan dibandingkan dengan suatu nilai (*r*) yang telah ditentukan. Apabila *random* angka tersebut lebih kecil dari nilai tersebut maka gen berisi nilai *False* yang artinya fitur tersebut tidak digunakan. Hal ini dilakukan pada tiap gen yang ada pada kromosom dalam satu populasi ketika generasi pertama. Misalkan $r = 0.3$

Gen 1	Gen 2	Gen 3	Gen 4	Gen 5	Gen 6	Gen 7	Gen 8	Gen 9	Gen 10
TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE

Gambar 3. Ilustrasi inisialisasi kromosom

Bilangan acak pada gen 1 ialah 0,35 yang nilainya lebih besar dari nilai r sehingga gen 1 berisi *True*. Sedangkan gen 2 ialah 0,20 yang nilainya lebih kecil dari nilai r sehingga gen 2 berisi nilai *False*.

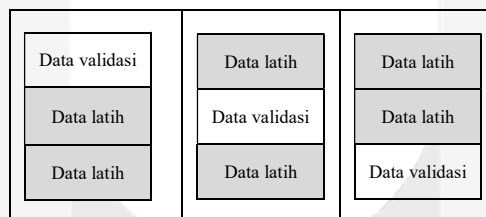
3.3.3 Evaluasi Fitness

Setiap Kromosom memiliki nilai fitness yang merupakan suatu ukuran baik tidaknya suatu solusi yang dinyatakan sebagai satu individu, atau dengan kata lain nilai fitness menyatakan nilai dari fungsi tujuan. Algoritma ini mempunyai tujuan untuk memaksimalkan nilai fitness atau mencari nilai fitness maksimal. Adapun untuk melakukan perhitungan fitness digunakan algoritma *Random Forest*.

Algoritma *Random Forest* merupakan metode yang bisa digunakan untuk klasifikasi maupun regresi. Cara kerjanya ialah dengan membangun *predictor* dengan sekumpulan pohon keputusan yang berkembang secara acak pada sub kumpulan *data training*. Hasil dari algoritma ini ialah modus dari pohon keputusan tersebut. Pada penelitian ini *Random Forest* bertujuan untuk menghitung akurasi yang nantinya digunakan sebagai nilai Fitness dalam proses GA. Adapun tahapan algoritma ini ialah :

1. Membagi data *training* dan testing sehingga $\frac{1}{3}$ dari data set sebagai data testing, sedangkan $\frac{2}{3}$ dari data set sebagai data *training*.
2. Menentukan nilai $k = 5, 15, \text{ dan } 20$ sebagai jumlah pohon yang ingin dibangun.
3. Membuat kumpulan sample acak dengan menarik contoh acak berukuran n dari kumpulan data *training*. Pembangunan data acak ini dilakukan hingga sejumlah pohon yang ingin dibangun.
4. Membuat pohon keputusan dari masing-masing sample acak tersebut
5. Melakukan *voting* dengan *majority vote* untuk setiap record pada data testing berdasarkan model yang dibuat pada setiap pohon keputusan. Hasil *vote* terbanyak akan menjadi nilai prediksi dari algoritma *Random Forest*.
6. Menghitung Akurasi

Dilakukan perhitungan akurasi menggunakan persamaan (1), dengan *Positive* ialah kelas 2, sedangkan *Negative* sebagai kelas 1. Perhitungan Akurasi pada Algoritma Genetika setelah melakukan *Random Forest* ialah dengan cara *Cross Validation*, metode statistik ini digunakan untuk mengevaluasi kinerja algoritma *Random Forest* dengan cara memisahkan data menjadi data latih dan data validasi. Pemilihan jenis CV dapat ditentukan berdasarkan ukuran dataset. Adapun untuk penelitian ini menggunakan 3 sehingga metode pengujian ini disebut *3 Fold Cross Validation*. Dilakukan pembagian secara acak menjadi 3 subset yang saling bebas. Lalu dilakukan pengulangan sebanyak 3 kali dengan pembagian 2 subset sebagai data latih dan 1 subset sebagai data validasi yang berbeda-beda. Tingkat akurasi dihitung dengan menggabungkan rata-rata ukuran kecocokan dengan melihat rata-rata pada setiap putaran untuk mendapatkan perkiraan kinerja model prediksi yang lebih akurat.



Gambar 4. Ilustrasi 3 Fold Cross Validation

3.3.4 Seleksi Populasi

Pada tahap ini, dilakukan seleksi orang tua untuk dimasukkan ke dalam populasi yang nantinya akan dilakukan proses reproduksi.

- Populasi Elitisme
Populasi Elitisme berisi kumpulan kromosom dengan nilai fitness tertinggi yang jumlahnya sudah ditentukan.
- Populasi Terbaik (*Best*)
Populasi *Best* atau Terbaik berisi kumpulan kromosom dengan nilai fitness tertinggi yang jumlahnya sudah ditentukan.
- Populasi Acak (*Random*)
Populasi *Random* berisi kumpulan kromosom yang dipilih secara acak dari Populasi yang telah diinisialisasi.

Penelitian ini dilakukan GA berjenis *generational replacement* yaitu, terdapat populasi Elitisme yaitu populasi yang diduplikat dan dijaga supaya tidak rusak oleh proses reproduksi. Hal ini bertujuan untuk menjaga agar

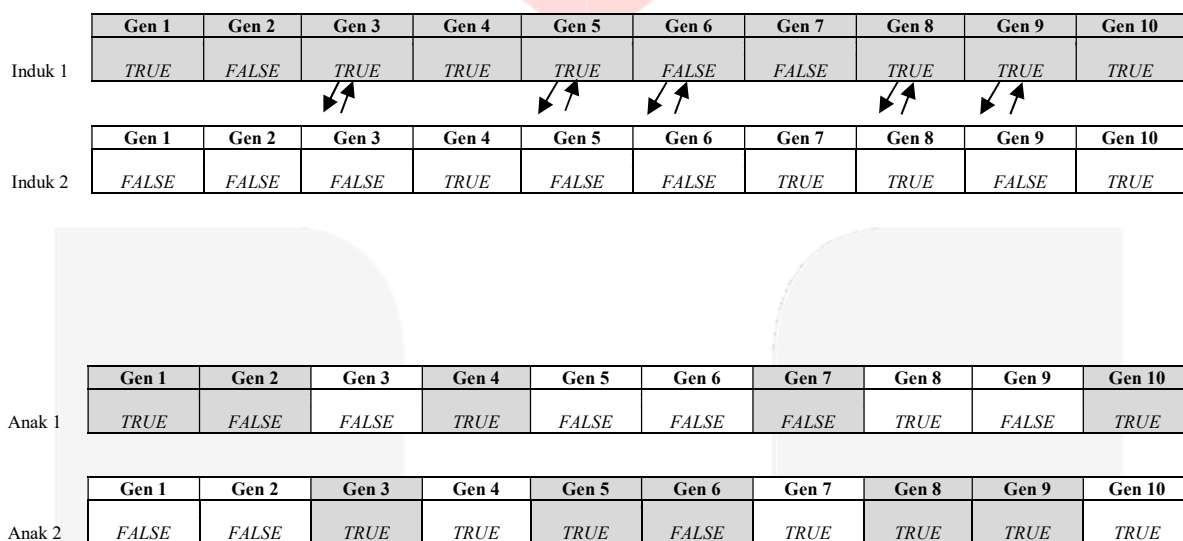
individu bernilai fitness tertinggi tidak hilang selama proses evolusi. Sedangkan untuk populasi best dan populasi *random* akan digabungkan untuk menjadi orang tua dan melanjutkan ke tahap selanjutnya.

3.3.5 Reproduksi Populasi

Proses ini bertujuan untuk menghasilkan populasi baru yang diharapkan lebih baik dari populasi sebelumnya.

- *Crossover*

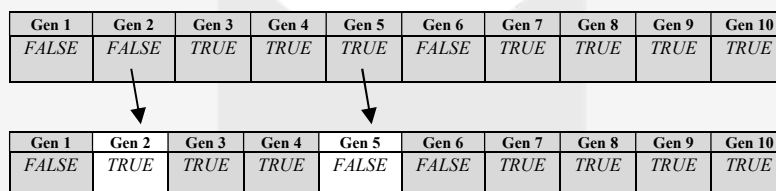
Proses pindah silang ini melibatkan dua individu sebagai orangtua dan akan menghasilkan anak dari individu baru. Prinsipnya ialah menentukan titik potong (posisi gen) secara *random*, yang digunakan sebagai acuan menukar gen-gen orang tua sehingga menghasilkan gen-gen anak. Jika bilangan *random* yang dibangkitkan (r) lebih kecil dari probabilitas *crossover* (pc) yaitu ($r < pc$) maka proses pindah silang dikerjakan. Terdapat beberapa jenis persilangan, dalam kasus ini digunakan persilangan banyak titik (*multi-point-crossover*) yaitu terdapat beberapa titik yang dijadikan sebagai acuan lokasi persilangan. Setiap pasangan kromosom menghasilkan 2 anak atau 2 kromosom baru.



Gambar 5. Ilustrasi proses *crossover*

- Mutasi

Pada proses ini, dilakukan pergantian gen dengan menukar nilai *True* menjadi *False* dan sebaliknya. Dilakukan pembangkitan bilangan *random* (r) pada kromosom untuk dibandingkan dengan Probabilitas mutasi (pm). Jika lebih kecil ($r < pm$) maka dilakukan proses mutasi. Setiap kromosom akan menghasilkan satu kromosom baru jika dilakukan mutasi.



Gambar 6. Ilustrasi proses mutasi

3.3.6 Pergantian Populasi

Proses ini mengganti populasi lama dengan gabungan populasi baru dari hasil reproduksi dan elitisme. Jumlah populasi lama dan populasi baru haruslah sama. Cara menggantinya ialah dengan mengambil populasi terbaik dari gabungan populasi elitisme dan populasi hasil reproduksi. Populasi ini akan menjadi solusi ketika telah sampai pada jumlah generasi atau siklus evolusi yang telah ditentukan. Jika belum, maka akan kembali pada proses evaluasi fitness.

4 Evaluasi

Pada bagian ini memaparkan pengujian dan analisis yang telah dilakukan sesuai dengan tujuan penelitian yang telah disampaikan pada bagian pendahuluan

4.1 Skenario

Adapun skenario yang dilakukan pada penelitian ini ialah:

4.1.1 Skenario 1 : Parameter GA

Pada Algoritma *Genetic Algorithm* terdapat beberapa parameter yang nilainya dapat disesuaikan. Pada penelitian ini, parameter-parameter tersebut ialah jumlah generasi, ukuran individu, probabilitas mutase (Pm), dan probabilitas crossover (Pc). Berdasarkan referensi [22] strategi membangun GA yang baik ialah menggunakan rentang nilai probabilitas *crossover* antara 0,5 sampai 0,95 sedangkan probabilitas mutasinya ialah rentang nilai antara 0,01 sampai 0,3. Oleh karena itu, dilakukan beberapa skenario dengan parameter GA pada table 6.

Tabel 7. Skenario Parameter GA

Generasi	Ukuran Individu	Pm	Pc
4	10	0,3	0,5
			0,7
			0,95
		0,1	0,5
			0,7
			0,95
		0,01	0,5
			0,7
			0,95
10	4	0,3	0,5
			0,7
			0,95
		0,1	0,5
			0,7
			0,95
		0,01	0,5
			0,7
			0,95

4.1.2 Skenario 2 : Jumlah Pohon pada *Random Forest*

Pada Algoritma *Random Forest* terdapat parameter jumlah pohon yang nilainya dapat disesuaikan. Oleh karena itu, dilakukan 3 percobaan dengan jumlah pohon yang berbeda.

Tabel 8. Skenario Jumlah Pohon

No	Jumlah Pohon
1	5
2	15
3	20

4.1.3 Skenario 3 : Data Set

Data yang tersedia ialah 7 Kumpulan Data berdasarkan regional. Pada penelitian untuk mengetahui hasil perbandingan kedua metode ialah scenario menggunakan data Regional 1 dan data Regional 2. Regional 1 meliputi kecamatan Ciwidey, Kutawaringin, Pasirjambu, Rancabali dan Soreang. Sedangkan Regional 2 meliputi Kecamatan Dateuhkolot, Katapang, Margaasih, dan Margahayu.

Tabel 9. Skenario Data Set

No	Data Set
1	Regional 1
2	Regional 2

42 Hasil Pengujian dan Analisis

4.2.1 Skenario 1

Berikut hasil dari pengujian menggunakan set data regional 1, dengan skenario Parameter GA:

Tabel 10. Hasil Pengujian skenario parameter GA pada data set Regional 1

Percobaan ke-	Generasi	Individu	Pm	Pc	Akurasi CV Terbaik (%)	Akurasi CV Rata-rata (%)	Akurasi CV sebelum GA (%)	Akurasi CV Sesudah GA (%)
1	4	10	0,01	0.5	89.44	87.46	83.06	89.44
2				0.7	89.44	86.5	83.06	89.44
3				0.95	91.53	90.17	83.06	91.53
4			0,1	0.5	89.44	87.54	83.06	89.44
5				0.7	87.36	85.54	83.06	87.36
6				0.95	91.53	89.83	83.06	91.53
7			0,3	0.5	89.44	86.79	83.06	89.44
8				0.7	89.44	87.458	83.06	89.44
9				0.95	91.53	87.56	83.06	91.53
10	10	4	0,01	0.5	89.44	89	83.06	89.44
11				0.7	91.53	87.68	83.06	91.53
12				0.95	89.44	89	83.06	89.44
13			0,1	0.5	89.44	90.39	83.06	91.53
14				0.7	91.53	90.30	83.06	91.53
15				0.95	91.53	90.1	83.06	91.53
16			0,3	0.5	91.53	88.35	83.06	91.53
17				0.7	89.44	88.16	83.06	89.44
18				0.95	91.53	90.43	83.06	91.53

Akurasi *Cross Validation*(CV) terbaik ialah nilai akurasi terbaik diantara banyak kromosom pada suatu percobaan, Akurasi CV Rata-rata ialah nilai akurasi rata-rata kromosom pada satu percobaan, Akurasi CV sebelum GA ialah nilai akurasi pada satu percobaan sebelum dilakukan proses GA. Serta Akurasi CV sesudah GA ialah nilai akurasi pada satu percobaan sesudah dilakukan proses GA

Pada Tabel 9, Akurasi CV sebelum GA selalu bernilai sama. Hal ini telah sesuai karena pada skenario ini hanya dilakukan perubahan pada parameter GA. Ketika data tersebut tidak mengalami proses GA maka tidak akan terjadi suatu perbedaan antar percobaan.

Terjadi peningkatan nilai dari Akurasi CV sebelum GA menuju Akurasi sesudah GA. Hal ini menunjukkan bahwa seleksi fitur yang dilakukan berhasil meningkatkan akurasi dengan mengurangi atribut yang kurang penting.

Nilai akurasi CV tertinggi pada setiap percobaan dengan nilai Pm dan Pc yang berbeda menunjukkan bahwa sulit menentukan nilai scenario Pm dan Pc terbaik karena tergantung pada bilangan acak yang dibangkitkan pada setiap kromosom ketika proses reproduksi.

Tidak terjadi perbedaan yang signifikan pada percobaan dengan jumlah generasi dan individu yang telah ditentukan. Pada 10 Generasi dengan 4 Individu menghasilkan sebanyak 40 individu dalam satu populasi, begitu pula ketika 4 Generasi dengan 10 Individu. Sehingga, hasil kromosom pada proses GA tidak jauh berbeda.

4.2.2 Skenario 2

Berikut hasil dari pengujian menggunakan data regional 1, dengan skenario Jumlah Pohon pada *Random Forest*:

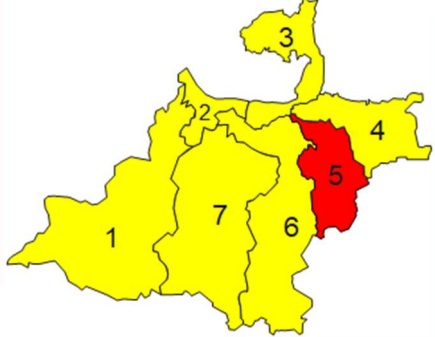
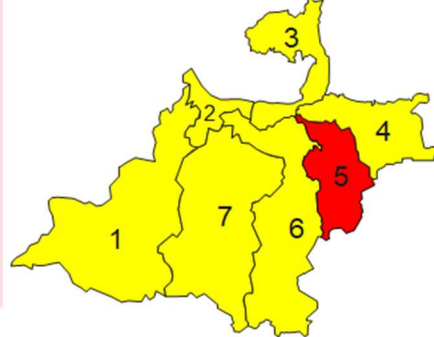
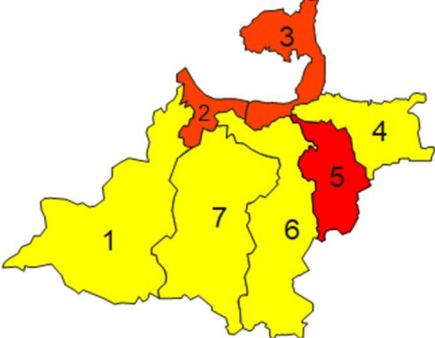
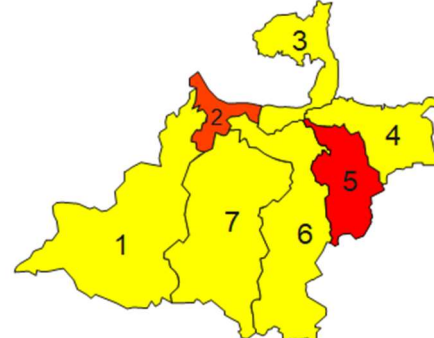

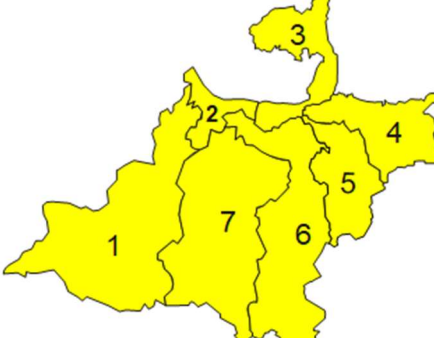
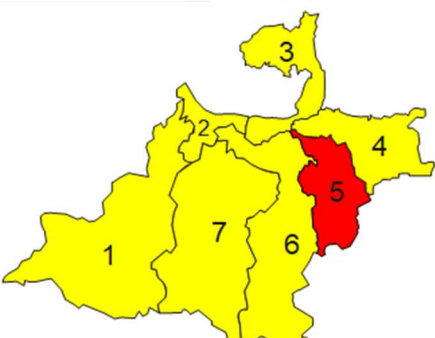
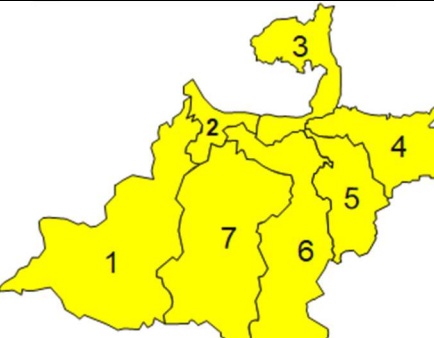
Tabel 11. Hasil Pengujian skenario jumlah pohon pada *Random Forest*

No	Pohon RF	Akurasi CV Terbaik	Akurasi CV Rata-rata	Akurasi CV sebelum GA	Akurasi CV Sesudah GA
1	5	89	87.43	87.22	89
2	15	91.53	89.9666	83.05556	91.53
3	20	91.53	90.43	83.06	91.53

AGS '14	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SEPT '14	1	1	1	1	1	1	1	1	1	1	1	1	1	1
OKT '14	1	1	1	1	1	1	1	1	2	2	1	1	1	1
NOV '14	1	1	1	1	1	1	1	1	2	2	1	1	1	1
DES '14	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JAN '15	1	1	1	1	1	1	1	1	1	1	1	1	1	1
FEB '15	1	1	2	2	2	2	1	1	2	2	1	1	1	1
MAR '15	1	1	2	2	2	2	1	1	2	2	1	1	1	1
APR '15	1	1	2	1	2	1	1	1	2	1	1	1	1	1
MEI '15	1	1	1	1	1	1	1	1	2	1	1	1	1	1
JUN '15	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JUL '15	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AGS '15	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SEPT '15	1	1	1	1	1	1	1	1	1	1	1	1	1	1
OKT '15	1	1	1	2	1	2	1	1	1	2	1	1	1	1
NOV '15	1	1	2	2	2	1	1	1	2	2	1	1	1	1
DES '15	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JAN '16	1	1	1	1	1	1	1	1	1	1	1	1	1	1
FEB '16	1	2	1	2	1	2	1	2	1	2	1	2	1	2
MAR '16	1	2	1	2	1	2	1	2	2	2	1	2	1	2
APR '16	1	1	1	1	1	1	1	1	2	1	1	1	1	1
MEI '16	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JUN '16	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JUL '16	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AGS '16	1	2	1	2	1	2	1	2	2	2	1	2	1	2
SEPT '16	1	2	2	2	2	2	1	2	1	2	1	2	1	2
OKT '16	1	2	1	2	1	2	1	2	1	2	1	2	1	2
NOV '16	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DES '16	1	1	1	1	1	1	1	1	2	1	1	1	1	1
JAN '17	1	1	1	1	1	1	1	1	2	1	1	1	1	1
FEB '17	1	1	1	2	1	1	1	1	2	2	1	1	1	1
MAR '17	1	1	1	1	1	1	1	1	1	1	1	1	1	1
APR '17	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MEI '17	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JUN '17	1	1	1	1	1	1	1	1	2	1	1	1	1	1
JUL '17	1	1	1	1	1	1	1	1	2	1	1	1	1	1
AGS '17	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SEPT '17	1	1	1	1	1	1	1	1	2	1	1	1	1	1
OKT '17	1	1	2	2	1	2	1	1	1	2	1	1	1	1
NOV '17	1	1	1	1	1	1	1	1	2	1	1	1	1	1

Beberapa hasil dari percobaan tersebut pada setiap bulan November dipetakan pada tabel 14.

Tabel 14. Geographic Information System Penyebaran Penyakit Demam Berdarah Dengue Kabupaten Bandung

Waktu	Hasil Prediksi	Data Asli
November 2014		
November 2015		
November 2016		
November 2017		

Tabel 14. Merupakan kumpulan dari hasil prediksi penyebaran penyakit DBD di Kabupaten Bandung yang terbagi dalam 7 Regional sesuai pada keterangan nomor. Warna kuning berarti penyebaran penyakit DBD pada daerah tersebut termasuk dalam kelas rendah (1) sedangkan warna merah artinya penyebaran penyakit DBD termasuk tinggi (2). Pembagian kelas pada data asli sesuai dengan rumus nomor 3 dan tercantum pada tabel nomor 5. Sedangkan pembagian kelas pada hasil prediksi berdasarkan metode *Random Forest* dengan 6 atribut yang terpilih dari proses GA.

5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa, Algoritma *Random Forest* dan Algoritma Genetika dapat diimplementasikan dalam prediksi penyakit demam berdarah *dengue* (DBD) di Kabupaten karena memiliki akurasi yang tinggi. Perbandingan dari hasil prediksi antara menggunakan Algoritma *Random Forest* (RF) tanpa *Genetic Algorithm* (GA) dan Hybrid Algoritma *Random Forest* dengan *Genetic Algorithm* menghasilkan rata-rata Akurasi 83,42% dan 89,12 %. Dari metode Hybrid Algoritma *Random Forest* dan *Genetic Algorithm* (GA) tersebut dilakukan seleksi fitur yang menunjukkan bahwa terdapat 6 atribut paling penting yang dapat digunakan untuk melakukan prediksi penyebaran penyakit DBD di Kabupaten Bandung, yaitu lama penyinaran matahari, kelembapan, *temperature*, curah hujan, jumlah kasus dan *Incident Rate*.

Daftar Pustaka

- [1] Kementerian Kesehatan RI, Sekretariat Jendral.2016.Profil kesehatan Tahun 2016. Jakarta: Kementerian Kesehatan RI
- [2] Kementerian Kesehatan RI, Sekretariat Jendral. 2015.Profil Kesehatan Indonesia Tahun 2015. Jakarta: Kementerian Kesehatan RI.
- [3] Dinas Kesehatan.2018.Profil Kesehatan Kabupaten Bandung Tahun 2018. Bandung : Dinas Kesehatan Kabupaten Bandung.
- [4] Surat Keputusan KPU RI NOMOR 275/PL.01.03-Kpt/06/KPU/IV/2018
- [5] Balagatabi, Zahra dkk. (2015). Comparison of Decision Tree Methods in Classification of Researcher's Cognitive styles in Academic Environment. Malaysia:Universitas Teknologi Malaysia
- [6] Lesnusa, Yopi Andry dkk. 2017. Aplikasi Jaringan Saraf Tiruan Backpropagation untuk Penyebaran Penyakit Demam Berdarah Dengue (DBD) di Kota Ambon
- [7] KESMAS. 2007engaruh Iklim Terhadap Kasus Demam Berdarah Dengue. Jurnal Kesehatan Masyarakat Nasional
- [8] Muliansyah & Tri baskoro (2016, April). Analisis Pola Sebaran Demam Berdarah Dengue Terhadap Penggunaan Lahan dengan Pendekatan Spasial di Kabupaten Banggai Provinsi Sulawesi Tengah Tahun 2011-2013
- [9] Muzakki, M. M., & Nhita, F. (2018, May). The Spreading Prediction of Dengue Hemorrhagic Fever (DHF) in Bandung Regency Using K-Means Clustering and Support Vector Machine Algorithm. In 2018 6th International Conference on Information and Communication Technology (ICoICT) (pp. 453-458). IEEE.
- [10] Nurul Chafid & Ismail Ari Wiboso. (2018, May). Implementasi Data Mining Untuk Clustering Daerah Penyebaran Penyakit Demam Berdarah di Kota Tangerang Selatan Menggunakan Algoritma K-Means
- [11] Breiman, Leo.2001. *Random Forests*. Netherlands: Kluwer Academic Publishers
- [12] Khan, S., Ullah, R., Khan, A., Sohail, A., Wahab, N., Bilal, M., & Ahmed, M. (2017). *Random Forest*-based Evaluation of Raman Spectroscopy for Dengue Fever Analysis. *Applied spectroscopy*, 71(9), 2111-2117.
- [13] Liu, Miao dkk. (2012). Comparison of *Random Forest*, Support Vector Machine and Back Propagation Neural Network for Electronic Tongue Data Classification: Application on the Recognition of Orange Beverage and Chinese Vinegar.
- [14] Budiharto, Widodo & Darwin Suhartono.2014. Artificial Intelligence. Yogyakarta:Andi Offset Hs, Widjono.2007.Bhs Ind Mt Kulh Pngemb Kepri DiPT (Rev).Jakarta:PT.Grasindo
- [15] Andriana, Arry. 2016. Optimisasi Rute Perjalanan dengan Algoritma Genetika
- [16] Yakoubi, Mohamed Amina, Mohamed tayeb Laskri. 2016. The Path Planning of Cleaner Robot for Coverage Region using *Genetic Algorithms*
- [17] Ratnawati, Dian Eka dkk. 2017. Prediksi Waktu Panen Tebu Menggunakan Gabungan Metode Backpropagation dan Algoritma Genetika
- [18] Almira, Lisa. 2018. Penggunaan Algoritma Genetika sebagai Seleksi Fitur dalam Pengenalan Ucapan pada Penderita Dysarthria
- [19] Wati, Risa. 2016. Penerapan Algoritma Genetika untuk Seleksi Fitur pada Analisis Sentimen Review Jasa Maskapai penerbangan Menggunakan Naïve Bayes
- [20] Wulan, Sisma Tri dkk. Optimasi Seleksi Fitur Klasifikasi Naïve Bayes Menggunakan Algoritma Genetika untuk Prediksi Risiko Kredit Konsumen
- [21] Sunge, Aswan Supriyadi. 2018. Optimasi Algoritma C4.5 Menggunakan Genetic Algoritma dalam Memprediksi Website Phishing
- [22] Basuki, Achmad. 2003b. Strategi Menggunakan Algoritma Genetika. [https:// lecturer.eepis-its.edu/~basuki/lecture/StrategiAlgoritmaGenetika.pdf](https://lecturer.eepis-its.edu/~basuki/lecture/StrategiAlgoritmaGenetika.pdf). Diakses Mei 2014.