

Identifikasi Status Bencana Gunung Berapi Menggunakan Metode Kombinasi Case-Based Reasoning dan Bayesian Network

Miftah Muhammad¹, Assaf Arief²

¹Program Studi Teknik Elektro

²Program Studi Teknik Informatika

Fakultas Teknik Universitas Khairun

Ternate, Indonesia

email: miftahmuh@unkhair.ac.id, assaf.arief@ui.ac.id

Abstrak— Secara geografis Indonesia berada di pertemuan dua lempeng yaitu Asia dan Australia, hal ini menyebabkan banyak terdapat gunung merapi. Selain itu pegunungan di Indonesia didominasi oleh gunung yang aktif dan berpotensi meletus sewaktu-waktu. Kondisi tersebut mengharuskan warga negara Indonesia, khususnya yang tinggal di sekitaran pegunungan berapi aktif tanggap akan bencana letusan gunung berapi. Pesatnya perkembangan Teknologi Informasi dan Komunikasi (TIK) pada mitigasi bencana dalam bentuk aplikasi pendeteksi potensi bencana gunung merapi, dapat menjadi solusi membantu warga masyarakat daerah rawan bencana. Penelitian ini menghasilkan suatu aplikasi identifikasi status gunung berapi. Data latih yang digunakan diambil dari situs Pusat Vulkanologi dan Mitigasi Bencana Gunung Berapi yang diklasifikasikan dalam tiga status yaitu normal, siaga, dan waspada. Pengklasifikasi menggunakan *Case-Based Reasoning (CBR)* dan *Bayesian Network (BN)* yang merupakan metode kombinasi berbasis probabilitas yang sederhana namun handal untuk meningkatkan akurasi data. Berdasarkan hasil pengujian dengan jumlah data latih terbesar mencapai tingkat akurasi 80%. Dengan demikian dapat dikatakan bahwa aplikasi identifikasi status bencana gunung berapi dengan metode CBR dan BN memiliki performa dan akurasi yang tinggi dalam mengklasifikasi status gunung berapi.

Kata Kunci; CBR; BN; status gunung berapi

I. PENDAHULUAN

Secara geografis Indonesia berada di pertemuan dua lempeng yaitu Asia dan Australia, sehingga di dominasi oleh gunung api yang terbentuk akibat zona subduksi antara kedua lempeng Eurasia dan lempeng Indo-Australia. Hingga tahun 2014 tercatat 127 gunung aktif yang dinyatakan masih aktif di Indonesia, dengan 5 juta warga berdiam di daerah gunung masih aktif tersebut. Selain itu, dari 127 gunung berapi aktif hanya 69 gunung yang terpantau itu artinya masih sekitar 40% lebih gunung aktif yang belum terpantau dengan baik. Dengan begitu, resiko bencana gunung berapi ketika terjadi letusan

gunung dampaknya kepada masyarakat masih sangat besar, mengingat masih banyak gunung yang belum terpantau baik dan jumlah warga yang menetap di daerah gunung berapi yang masih aktif begitu banyak.

Dengan potensi bencana alam yang besar, sudah selayaknya Indonesia mempunyai pengelolaan bencana yang konsisten, efektif dan terpadu, baik dalam tingkat pemerintahan pusat maupun dalam tingkat pemerintahan daerah. Kemajuan teknologi informasi dan komunikasi (TIK) yang mendukung aktivitas pengelolaan bencana alam dapat diterapkan dalam upaya untuk mengoptimalkan proses penanggulangan bencana dan proses pemulihan wilayah setelah terjadi bencana alam. Pengelolaan bencana adalah suatu proses yang berkelanjutan, bukan proses sesaat, mengingat bencana alam pasti akan terjadi. Pengelolaan bencana mutlak diperlukan dengan tujuan utama adalah peningkatan kepedulian semua pihak dalam upaya untuk mengurangi dampak bencana alam. Oleh karena itu, informasi mengenai wilayah yang terkena bencana dan identifikasi besarnya kerusakan sangat diperlukan sebagai salah satu kebijakan pengelolaan bencana dalam upaya mengurangi dampak bencana.

Penelitian-penelitian dalam dunia *artificial intelligence* untuk gunung berapi pernah dilakukan oleh beberapa peneliti sebelumnya diantaranya [1] dan [2]. Pada penelitian [1] dengan menggunakan *naïve bayes* sedangkan penelitian dari [2] menggunakan metode ANFIS. Penelitian ini memanfaatkan data-data lama untuk proses memberikan identifikasi status gunung berapi atau yang dikenal dengan *Case-Based Reasoning*. Setelah menggunakan metode CBR di lanjutkan dengan metode *Bayesian Network* untuk proses penalaran telah di uji pada beberapa kasus-kasus lain dengan akurasi sistem bisa mencapai lebih dari 90% bahkan bisa 100%. Sebagaimana yang diteliti oleh [3] membuat sistem pendukung keputusan dengan menggunakan metode CBR pada diagnosis paru akibat kerja dengan akurasi sistem mencapai 95, 3%. Selain itu, pada tahapan retrieve dari CBR digunakan Bayesian Network agar secara probabilitas teruji dengan data yang terukur.

Proses identifikasi status gunung berapi ada beberapa faktor kegunaan yang dilihat agar para pakar bisa mengambil kesimpulan terkait dengan aktivitas dari setiap gunung berapi. Pada penelitian ini ada empat faktor kegunaan yang dipakai, yaitu gempa vulkanik dangkal, tektonik jauh, gempa hembusan dan vulkanik dalam. Sedangkan fitur yang dijadikan sebagai kriteria untuk rekomendasi status gunung ada lima fitur yaitu empat jenis kegunaan dan status sebelum dari gunung tersebut. Sedangkan untuk status yang direkomendasikan ada tiga status yaitu status normal, waspada, dan siaga.

II. TINJAUAN PUSTAKA

A. Case-Based Reasoning (CBR)

CBR telah diimplementasikan dalam berbagai bidang yang berbeda dan telah teruji yaitu, kedokteran, ekonomi, hukum, komputasi, jaringan komunikasi, desain pabrik, keuangan, penjadwalan, bahasa, sejarah, makanan/nutrisi, penemuan rute dan lingkungan [4].

Inti dari CBR adalah memecahkan masalah berdasarkan data-data lampau. Selain itu pengetahuan dari CBR juga bersifat dinamis karena selalu ada penambahan pengetahuan pada case base. CBR juga dapat diorientasikan sebagai suatu siklus proses seperti terlihat pada gambar 1, yang dibagi menjadi empat sub proses [5], yaitu:

- Retrieve: Mengambil kasus yang paling menyerupai/relevan (*similar*) dengan kasus yang baru.
- Reuse: Menggunakan kembali pengetahuan dan informasi dalam kasus untuk memecahkan masalah
- Revise : Merevisi solusi yang diusulkan
- Retain: Menyimpan pengalaman untuk tujuan di masa depan pada kasus-kasus yang lain.

B. Bayesian Network (BN)

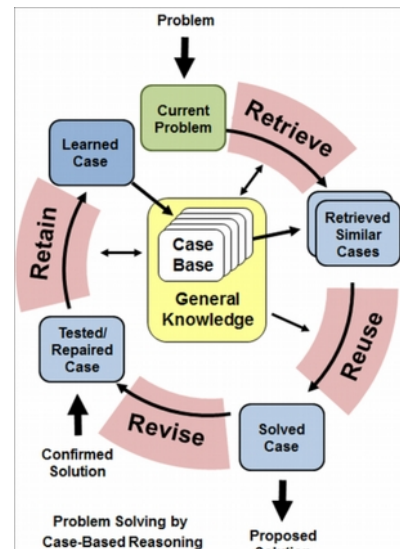
Bayesian Networks (BN) merupakan representasi grafis dari distribusi probabilitas, yaitu berupa *Directed Acyclic Graph* (DAG) yang terdiri dari satu set *node* untuk mewakili variabel dan satu set *link* diarahkan untuk menghubungkan pasang *node*. Setiap *node* memiliki distribusi probabilitas bersyarat yang mengkuantifikasi hubungan probabilistik antara *node* dan induknya sehingga untuk jaringan *node*.

Variabel G bisa ditentukan probabilistik dengan memanfaatkan pengetahuan dari variabel yang telah diketahui. Di dalam struktur Bayesian Network ini disebut dengan *inferensi probabilistik*, yaitu memanfaatkan pengetahuan tentang keterhubungan diantara variabel yang pada graf Bayesian Network untuk melakukan pengambilan keputusan (*inferensi*) probabilistik. Berdasarkan graf di atas, probabilistik dari variabel D dapat dicari jika diketahui nilai C.

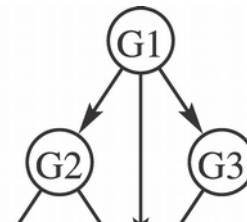
Untuk melakukan inferensi probabilistik terlebih dahulu menghitung probabilistik gabungan Bayesian Network dengan:

$$P(X | \theta) = \prod_{i=1}^n p(x_i | pa(x_i), \theta_i) \quad (1)$$

n adalah banyaknya variabel dan $pa()$ adalah *parent* dari variabel x ke-i dan adalah parameter ke-i di mana i dimulai dari 1 sampai banyaknya variabel pada graf.



Gambar 1. Siklus CBR



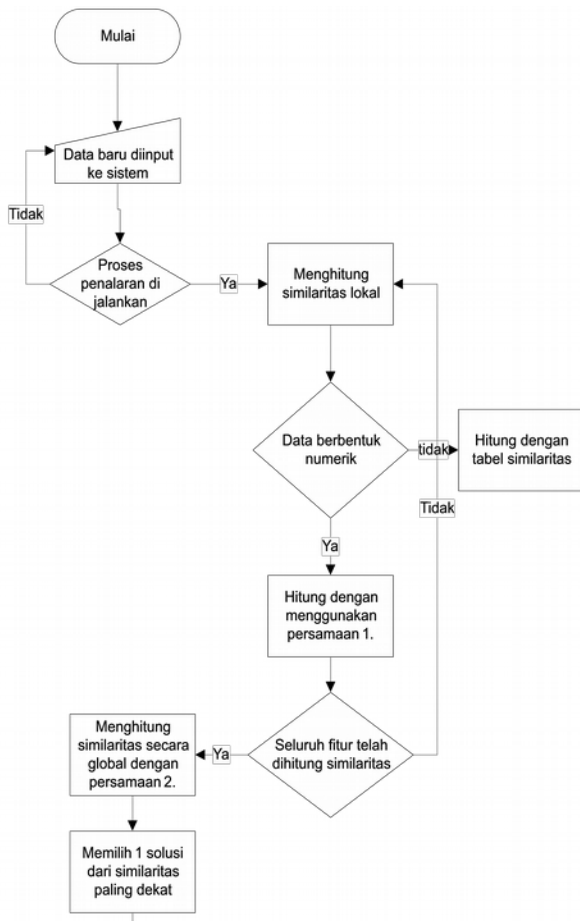
Gambar 2. Graf Bayesian Network

III. METODE PENELITIAN

Penelitian ini mencoba membangun suatu pendekatan algoritma *data mining* yaitu CBR dan data-data tersebut di proses dengan BN[6]. Proses CBR secara singkatnya dapat dirangkum antara lain mengidentifikasi suatu permasalahan pada suatu kasus, mencari kasus di masa lalu yang memiliki kemiripan gejala dengan kasus baru, mengambil solusi pada kasus lama untuk dievaluasi, mengadaptasikan solusi terhadap kasus baru dan memperbaharui sistem yang telah dibangun sebelumnya dengan memasukkan kasus baru beserta solusinya ke dalam sistem.

A. Proses Retrieval

Proses retrieval adalah proses untuk mengambil kasus kasus yang serupa atau kasus-kasus yang paling menyerupai antara kasus baru dengan kasus yang disimpan di case base. Supaya efektif harus ada suatu kriteria yang menentukan bagaimana cara menelusuri dan mengontrol suatu kasus di dalam case base. Sebagian besar proses penelusuran dilakukan dengan mencari keseluruhan case based dengan membandingkan isi features dengan kasus yang baru [8]. Untuk mengambil kasus yang serupa dilakukan tahapan perhitungan similaritas pada penelitian ini similaritas yang dihitung dibagi dua yaitu similaritas local dan global. Pada similaritas lokal ada dua tipe data yaitu data berbentuk numeric dan data berbentuk kategori. Tahap-tahap proses retrieval seperti ditunjukkan pada gambar 3.

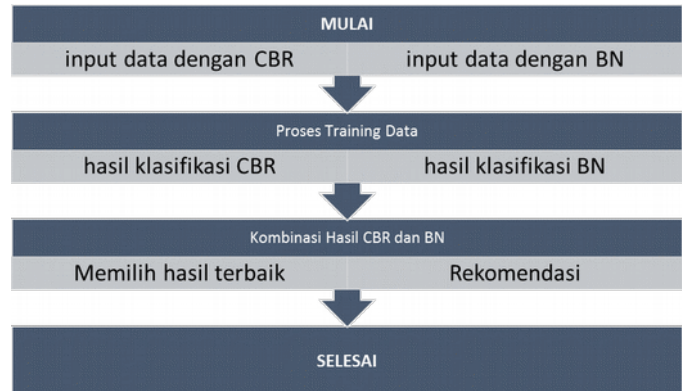


Gambar 3. Proses retrieval

B. Metode Kombinasi CBR dan BN

Pada proses selanjutnya adalah penggabungan secara paralel metode CBR dan BN secara bersamaan. Proses input data gunung berapi yang diambil dari data Pusat Vulkanologi dan Mitigasi Bencana Gunung Berapi,

Kementerian Energi dan Sumber Daya Mineral. Adapun langkah-langkah penelitian seperti ditunjukkan pada gambar 4 dibawah ini.

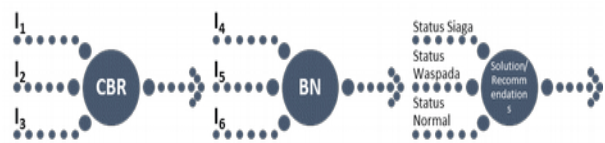


Gambar 4. Langkah-langkah penelitian

Adapun cara mengkombinasikan CBR dan BN dilakukan dengan metode sebagai berikut:

- Secara paralel
- Secara berurutan CBR-BN
- Secara berurutan BN-CBR

Kombinasi metode secara paralel, dilakukan dengan cara semua metode baik CBR dan BN menggunakan semua variabel input masing-masing dan kemudian menghasilkan klasifikasi secara independen. Hasilnya kemudian dibandingkan dengan pilihan yang terbaik, dan klasifikasi terbaik dipilih. Fokus kami adalah pendekatan yang terintegrasi, diwakili oleh dua kombinasi berurutan. BN dan CBR adalah terhubung sedemikian rupa sehingga sistem pertama dalam urutan menghitung sesuatu yang dibutuhkan sistem kedua. Proses tahapan kombinasi CBR dan BN seperti di tunjukan pada gambar 5. dibawah.



Gambar 5. Arsitektur kombinasi Metode CBR dan BN

IV. HASIL DAN PEMBAHASAN

Setelah melakukan fase penginputan data traning dengan CBR (I₁, I₂, I₃...dst) dengan proses pengukuran similaritas pada kasus-kasus baru maupun lama, sehingga yang menghasilkan output 1, output 2...dst. Selanjutnya dilakukan penginputan data (I₄, I₅, I₆..dst) dilakukan dengan metode BN sehingga menghasilkan solusi 1, solusi 2 dan Solusi 3. Setelah itu dilakukan penstausan dari hasil pengolahan data CBR dan BN sehingga dihasilkan solusi dengan akurasi yang lebih baik.

Pengukuran similaritas pada tahapan CBR akan menghasilkan nilai yang menentukan tentang ada atau

tidaknya kemiripan antara kasus yang baru dengan kasus-kasus yang ada dalam *case base*. Selain itu pengukuran similaritas juga untuk mendukung konsep dari algoritma *nearest neighbor* (NN) yaitu nilai dengan similaritas tertinggi yang dipilih sebagai tetangga terdekat dari kasus baru tersebut [9]. Dalam pengukuran similaritas dibagi menjadi dua kategori yaitu similaritas lokal dan global.

Similaritas lokal: Nilai similaritas lokal dibedakan menjadi dua jenis, yaitu similaritas lokal numerik dan simbolik. Pada fitur vulkanik dangkal, vulkanik dalam, tektonik jauh dan gempa hembusan menggunakan similaritas numerik dengan persamaan yaitu persamaan 1. Sedangkan untuk fitur status sebelumnya menggunakan similaritas tabel atau matriks seperti pada tabel 1 berikut:

$$f(S_i, T_i) = 1 - \frac{|S_i - T_i|}{f_{\max} - f_{\min}} \quad (2)$$

TABEL 1. PENGUKURAN SIMILARITAS LOKAL

S _i /T _i	St. Normal	St. Waspada	St. Siaga
St. Normal	1	0.6	0.2
St. Waspada	0.6	1	0.6
St. Siaga	0.2	0.6	1

- Similaritas global: digunakan untuk menghitung keserupaan antar masalah baru dengan kasus yang tersimpan dalam basis kasus, dengan persamaan yang digunakan yaitu persamaan 2 [10].

$$Sim(S, T) = \frac{\sum_{i=1}^n f(S_i, T_i)}{\sum_{i=1}^n 1} \quad (3)$$

Selanjutnya dalam CBR dilakukan pengukuran akurasi sistem yang dihitung dengan membandingkan jumlah rekomendasi benar pada status gunung berapi dengan jumlah data uji status gunung berapi perbandingannya dapat dituliskan dalam bentuk persamaan 4 berikut:

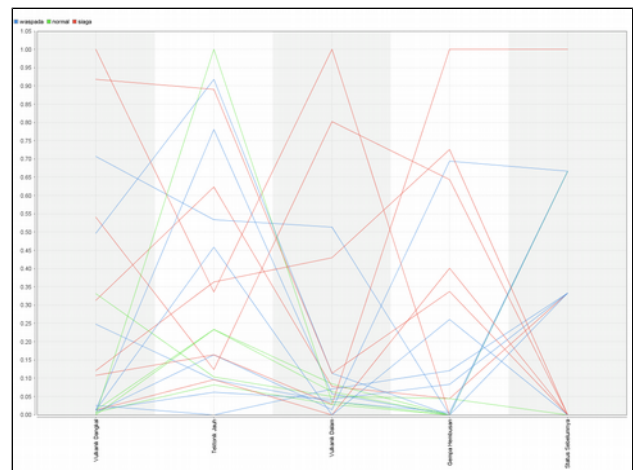
$$akurasi = \frac{rekomenasibenar}{\sum dataUji} \times 100\% \quad (4)$$

Dalam pengujian sistem kombinasi CBR dan BN untuk indentifikasi status gunung berapi menggunakan data-data aktivitas gunung berapi yang berada di Indonesia. data set tersebut diambil dari situs resmi Pusat Vulkanologi dan Mitigasi Bencana Gunung Berapi, Kementerian Energi dan Sumber Daya Mineral. Sebagaimana pada table 2 adalah 20 data yang disimpan di CBR.

TABEL 2. PROSES PERHITUNGAN DENGAN CBR

No	VD	TJ	VDA	GH	SS	SR
1	2	68	1	41	waspada	waspada
2	0	35	10	0	waspada	normal
3	19	54	62	114	waspada	siaga
4	2	10	7	13	Siaga	Waspada
5	2	15	1	63	waspada	siaga
6	144	131	17	53	waspada	siaga
7	78	135	17	0	normal	waspada
8	1	13	5	0	waspada	Normal
9	17	25	5	157	siaga	siaga
10	0	147	6	0	normal	normal
11	85	19	115	101	waspada	siaga
12	4	1	11	19	Siaga	waspada
13	2	115	9	0	waspada	waspada
14	157	50	143	0	waspada	siaga
15	111	79	74	0	Siaga	waspada
16	1	25	3	109	normal	waspada
17	52	16	8	7	waspada	normal
18	2	35	13	0	waspada	normal
19	39	15	6	1	normal	waspada
20	49	92	12	7	Siaga	siaga

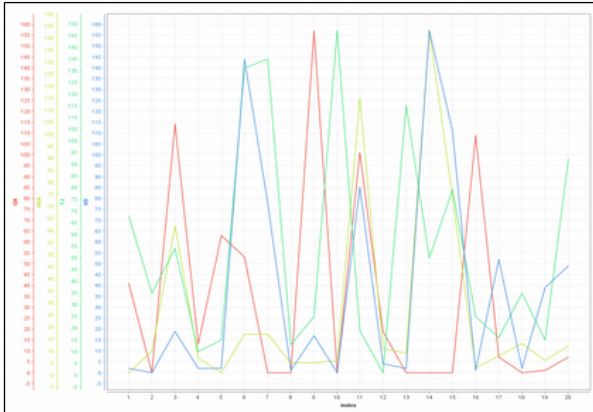
Keterangan singkatan:
 VD = Vulkanik Dangkal
 TJ = Tektonik Jauh
 VDA = Vulkanik Dalam
 GH = Gempa Hembusan
 SS = Status sebelumnya
 SR = Status yang direkomendasi



Gambar. 6. Tampilan Grafik Data dengan CBR

Selanjutnya hasil perhitungan dengan CBR menjadi input proses pada metode BN untuk dilakukan uji *cross-validation* kemudian dimasukkan training dataset dengan *Naïve Bayes* dan dilakukan *testing* dengan *performances classification*. Simulasi dataset *bencana_gunung.csv* yang

diambil dari input data CBR. Lalu dilakukan pengukuran akurasi sistem dengan data training dan testing menggunakan software data mining Rapid Miner Studio 2017, ditampilkan hasil keluaran seperti di tunjukan oleh gambar 7 berikut.



Gambar 7. Status rekomendasi hasil CBR dan BN

Proses pengujian data-data aktivitas pada gunung berapi dilakukan dengan tiga skenario. Skenario pertama diuji dengan menggunakan 3 fitur yaitu vulkanik dangkal, vulkanik dalam dan tektonik jauh dimana akurasi sistem hanya 55 % dengan data di *case base* sebanyak 30 data. Pada skenario kedua peneliti menambahkan 2 fitur lain yaitu gempa hembusan dan status gunung berapi sebelumnya sehingga jumlah fitur menjadi 5 fitur. dengan data uji sebanyak 30 data dan data di *case base* sebanyak 10 data, didapat akurasi sistem sebesar 76,66% atau ada 23 data keputusan sistem yang sesuai dengan data aslinya. Dan ada 7 data keputusan sistem yang tidak sesuai dengan data aslinya. Kemudian pada skenario ketiga, dengan 5 fitur dan 20 data digunakan sebagai input pada *case base* status sebelumnya dan didapat akurasi sistem sebesar 80% atau ada 24 data uji keputusan sistem sesuai dengan data aslinya. Sedangkan yang tidak sesuai ada 6 data. Perbandingan kemampuan sistem bisa dilihat pada tabel 3 berikut:

TABEL 3. PROSES PENGUJIAN DATASET

no.	model pengujian	akurasi (%)
1	Skenario 1	55%
2	Skenario 2	76,66 %
3	Skenario 3	80%

V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pengujian kombinasi metode *Case-Based Reasoning* (CBR) dan *Bayesian Network* (BN) pada preses pengidentifikasian status bencana gunung berapi maka dapat disimpulkan terdapat perbedaan akurasi sistem antara 6 fitur dan 10 fitur dan variabel, dan *case base* yang berjumlah 20 data dan 30

data. Sehingga untuk meningkatkan akurasi data membutuhkan banyak sampel dataset sebagai input lanjutan.

Selanjutnya untuk penelitian berikutnya diharapkan pengujian data aktivitas gunung berapi dilakukan hanya dengan data satu gunung, mengingat setiap gunung berapi memiliki cirri masing-masing yang berbeda. Dan penanganan akurasi prediksinya perlu digunakan algoritma lain yang dapat meningkatkan akurasi sistem penentuan status gunung berapi.

UCAPAN TERIMA KASIH

Pada penelitian ini kami ucapkan terimakasih kepada pihak kemenristekdikti yang telah membiayai penelitian ini hingga selesai. Kepada kepala dan pengurus laboratorium Rekayasa Perangkat Lunak (RPL) 2015-2017 Teknik Informatika Universitas Khairun yang memfasilitasi laboratorium kepada kami selama melakukan proses penelitian.

DAFTAR PUSTAKA

- [1] Lillian, D, Y., 2014, "Sistem Peringatan Dini Status Gunung Berapi Menggunakan Pengklasifikasi Naïve Bayes", 27 Februari-1 Maret 2014 [Proceeding KNSI Makassar, 835-838].
- [2] Fatkhurrozi, b., muslim, m, a., dan santoso, d, r., "penggunaan artificial neuro fuzzy inference sistem (anfis) dalam penentuan status aktivitas gunung merapi" vol. 6, no. 2, jurnal eccis.2012. 113-118.
- [3] Tomar, S. P. P., Singh, R., Saxena, K. P., dan Sharma, J., Case Based Medical Diagnosis of Occupational Chronic Lung Diseases From Their Symptoms and Signs, Vol 5, International Journal of Biometrics and Bioinformatics (IJBB), 2011, 216-224
- [4] Mulyana, S., dan Hartati, S., Tinjauan Singkat Perkembangan Case-Based Reasoning, 1979-2328, 23 Mei 2009 [Seminar Nasional Informatika (SEMNASIF), Yogyakarta, D17-D24]
- [5] Aamodt, A., dan Plaza, E., Case Based Reasoning: Foundation Issues Methodological Variations, and System Approaches, vol 7, AI Communication IOS Press, 1994, pp. 39-59.
- [6] Wu, X., and Kumar, V., The Top Ten Algorithms in Data Mining. London: CRC Press Taylor & Francis Group, 2009.
- [7] Guessoum, S., Laskri, M. T., dan Lieber, J., A Case-Based Reasoning System for the Diagnosis of Chronic Obstructive Pulmonary Disease, Vol. 41 Expert Systems With Applications An International Journal. 2014, 267-273.
- [8] Pal, S, K., Shiu, S, C, K., Foundations of Soft Case-Based Reasoning. Wiley-Interscience publication, 2004
- [9] Shi, H., Xin, M., Dong, W., A Kind of Case Similarity Model Based on Case-Based Reasoning, *International Conferences on Internet of Things, and Cyber, Physical and Social Computing IEEE*, 2011, 453-457
- [10] Ong, L., S., Shepherd, B., Tong, L. C., Choen, F. S., Ho, H. K., Tang, C. L., Ho, Y. S., dan Tan, K., The Colortecal Cancer Recurrence Support (CARES) System, Vol 11, *Artificial intelligence in Medicine Elsevier*, 1997, 175-188.