Pengembangan Aplikasi Pendeteksi Potensi Bencana Gunung Berapi Menggunakan Pengklasifikasi Bayesian

Dewi Yanti Liliana

Prodi Teknik Informatika

Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta

Jalan Prof. G.E. Siwabessy, Kampus UI, Depok, Jawa Barat

dewiyanti.liliana@tik.pnj.ac.id

Diterima: 15 Maret 2015. Disetujui: 23 April 2015. Dipublikasikan: Mei 2015

Abstrak - Letak geografis Indonesia berada di pertemuan dua lempeng yaitu Asia dan Australia, menyebabkan banyak terdapat gunung berapi. 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 sekitar pegunungan berapi aktif tanggap akan bencana letusan gunung berapi. Berkembangnya sains dan teknologi informasi dapat memberi kontribusi dalam rangka mitigasi bencana geologi yang diakibatkan aktivitas vulkanik dalam bentuk aplikasi pendeteksi potensi bencana gunung berapi. Penelitian ini menghasilkan suatu aplikasi pendeteksi 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 Bayesian digunakan karena merupakan metode berbasis probabilitas yang sederhana namun handal. Berdasarkan pengujian jumlah data latih terbesar mencapai tingkat akurasi 90%. Dengan demikian dapat dikatakan bahwa aplikasi pendeteksi potensi bencana gunung berapi memiliki performa yang baik dalam mengklasifikasi status gunung berapi.

Kata Kunci: klasifikasi, bencana gunung berapi, pengklasifikasi bayesian, pembelajaran tersupervisi.

I. PENDAHULUAN

Posisi geografis Indonesia terletak di pertemuan dua lempeng Asia dan Australia, hal ini menyebabkan Indonesia sering mengalami gempa tektonik dan banyak muncul gunung berapi [1]. Indonesia memiliki lebih dari 400 gunung berapi dan 130 di antaranya termasuk gunung berapi aktif [2]. Sebagian dari gunung berapi terletak di dasar laut dan tidak terlihat dari permukaan laut. Indonesia merupakan tempat pertemuan dua rangkaian gunung berapi aktif (*ring of fire*). Terdapat puluhan patahan aktif di wilayah Indonesia. Menurut Kepala Pusat

Vulkanologi dan Bencana Geologi, Kementerian Energi dan Sumber Daya Mineral, dari 127 gunung api aktif di Indonesia hanya 69 yang terpantau dan itupun masih jauh dari keadaan ideal, baik dari segi peralatan maupun dari segi Sumber Daya Manusia [1].

Berdasarkan kondisi nyata yang ada di Indonesia maka disusunlah sebuah penelitian untuk mengimplementasikan aplikasi pendeteksi potensi bencana gunung berapi guna mengantisipasi dampak serta kerugian akibat bencana letusan gunung berapi. Status gunung berapi diklasifikasikan menjadi tiga kategori yaitu normal, siaga, dan waspada. Pengklasifikasi yang digunakan adalah Bayesian atau disebut *Naïve Bayes Classifier* (NBC), sebuah metode pembelajaran mesin sederhana yang handal untuk mengenali pola. NBC bekerja dengan mencari probabilitas posterior dari suatu pola, kemudian mengklasifikasikan pola tersebut ke dalam kelas yang memiliki nilai posterior terbesar [3].

NBC merupakan metode pembelajaran tersupervisi, oleh karena itu diperlukan data latih sebagai data pembelajaran yang diambil dari situs resmi Pusat Vulkanologi dan Mitigasi Bencana Gunung Berapi (proxy.vsi.esdm.go.id/index.php). Data-set tersebut akan digunakan sebagai vektor fitur untuk mengklasifikasikan status gunung berapi.

II. PERSAMAAN MATEMATIKA

A. Teorema Bayes

Teorema Bayes merupakan pendekatan statistika dan probabilistik yang mendasari bidang pengenalan pola [4]. Jika X melambangkan himpunan fitur atau data dan Y melambangkan variabel kelas, apabila variabel kelas Y memiliki hubungan non-deterministik dengan fitur X maka X dan Y dapat dinyatakan sebagai variabel acak yang memiliki hubungan peluang bersyarat $P(Y|X)$.

Peluang bersyarat $P(Y|X)$ juga dikenal dengan posterior probability. Bentuk umum persamaan Bayes adalah:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

dimana:

X : data atau fitur yang ingin diduga kelasnya

H : hipotesis bahwa X merupakan anggota kelas H

$P(X)$: probabilitas X

$P(H)$: probabilitas awal dari kelas H (*prior probability*)

$P(X|H)$: probabilitas X bersyarat kelas H

$P(H|X)$: probabilitas kelas H dengan kondisi data X (*posterior probability*)

$P(X|H)$: *conditional probability* yang disebut dengan *likelihood* atau kecenderungan kejadian X dengan syarat/kondisi kelas H .

Nilai *prior probability* dan *likelihood* dapat diperoleh dari eksperimen terhadap data latih [5]. Data X (dapat berupa vektor fitur) diklasifikasikan sebagai anggota kelas C jika *posterior probability* $P(C_i|X)$ adalah yang terbesar dibandingkan dengan *posterior probability* $P(C_k|X)$ dari seluruh kelas k .

Untuk data latih yang bersifat kontinu dapat digunakan persamaan 2 untuk mencari nilai *likelihood* (disebut juga formula distribusi Gauss) [1].

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp \left(-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right) \quad (2)$$

dimana P melambangkan probabilitas, X_i adalah fitur ke- i , Y merupakan kelas y_j , dan μ_j and σ_j^2 adalah rata-rata dan varian dari kelas j .

B. Pengklasifikasi Bayesian (NBC)

NBC adalah pengklasifikasi sederhana berbasis probabilitas yang menerapkan teorema Bayes. Keuntungan dari penggunaan NBC adalah hanya diperlukan relatif sedikit data latih untuk mengestimasi parameter yang diperlukan pada proses klasifikasi [6]. NBC mengestimasi probabilitas dari suatu kelas bersyarat dengan asumsi bahwa atribut data (fitur) adalah bebas bersyarat (*conditionally independent*) terhadap variabel kelas [7]. *Conditional independent* diekspresikan dalam persamaan 3, yaitu:

$$P(X|Y = y) = \prod_{i=1}^d P(X_i|Y = y) \quad (3)$$

dimana $X = \{X_1, X_2, \dots, X_d\}$ adalah banyaknya fitur (d buah fitur). Untuk mengklasifikasikan data tes, NBC

menghitung probabilitas posterior dari setiap kategori kelas Y lalu memutuskan kelas dengan nilai probabilitas posterior terbesar sebagai kategori kelas suatu data uji. Perhitungan probabilitas posterior $P(Y|X)$ suatu data latih yang memiliki beberapa fitur X_i menggunakan formula 4.

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^d P(X_i|Y)}{P(X)} \quad (4)$$

III. METODOLOGI PENELITIAN

Metodologi penelitian meliputi beberapa tahapan yang digambarkan dalam diagram alir pada Gambar 1.

Gambar 1. Diagram Alir Tahapan Penelitian

Proses penelitian diawali dengan pengumpulan data latih yang disebut dengan data set status gunung berapi. Selanjutnya ditentukan fitur-fitur yang *invariant*, yaitu fitur yang merepresentasikan karakteristik status gunung berapi. Tahap berikutnya yaitu pemilihan model, bagaimana memodelkan fitur-fitur gunung berapi untuk mencari parameter probabilitas posterior dengan metode NBC berdasarkan data set yang ada. Selanjutnya dilakukan pengujian performansi aplikasi yang dibangun. Terakhir dilakukan evaluasi

aplikasi untuk mengukur akurasi aplikasi yang telah dihasilkan.

A. Pengumpulan Data

Pada tahap pengumpulan data dikumpulkan data latih dari gunung berapi yang ada di Indonesia. Data-set ini berasal dari Pusat Vulkanologi dan Mitigasi Bencana Gunung Berapi, Kementerian Energi dan Sumber Daya Mineral. Data didapatkan pada situs resmi pusat tersebut, yaitu proxy.vsi.esdm.go.id/index.php [8]. Aktivitas suatu gunung berapi dipantau setiap hari oleh pusat tersebut sehingga reliabilitas data yang didapatkan bisa dipertanggung jawabkan.

B. Pemilihan Fitur

Berdasarkan pengamatan terhadap data yang dikumpulkan oleh Pusat Vulkanologi dan Mitigasi Bencana Gunung Berapi didapatkan beberapa data latih yang memiliki beberapa fitur dan kelas kategori[2]. Fitur tersebut adalah Vulkanik Dangkal, Tektonik Jauh, dan Vulkanik Dalam. Sedangkan Kategori kelas yaitu siaga, waspada, dan normal. Tabel 1 dan Tabel 2 menunjukkan data-set yang digunakan.

TABEL 1. DATA LATIH STATUS GUNUNG BERAPI

Data	Fitur			Keputusan
	Vulkanik Dangkal	Tektonik Jauh	Vulkanik Dalam	
Data 1	16	113	15	Siaga
Data 2	2	67	2	Waspada
Data 3	44	4	128	Siaga
Data 4	79	1	24	Waspada
Data 5	77	7	6	Waspada
Data 6	1	60	1	Waspada
Data 7	10	30	45	Siaga
Data 8	44	3	62	Normal
Data 9	143	2	3	Siaga
Data 10	49	2	1	Waspada
Data 11	1	49	8	Normal
Data 12	9	29	43	Normal
Data 13	2	130	24	Siaga
Data 14	1	40	3	Normal
Data 15	13	114	39	Siaga
Data 16	3	47	39	Normal
Data 17	51	3	7	Waspada
Data 18	45	13	93	Siaga
Data 19	6	51	77	Siaga
Data 20	13	59	27	Normal
Data 21	78	4	57	Waspada
Data 22	4	76	37	Normal
Data 23	4	76	85	Siaga
Data 24	65	70	6	Siaga
Data 25	70	45	34	Waspada
Data 26	17	90	6	Siaga
Data 27	8	22	92	Siaga
Data 28	92	26	10	Waspada
Data 29	5	78	10	Normal
Data 30	12	17	58	Siaga

C. Pemilihan Model

Tabel 2 menunjukkan data uji status gunung berapi.

TABEL 2. DATA UJI STATUS GUNUNG BERAPI

Data	Fitur		
	Vulkanik Dangkal	Tektonik Jauh	Vulkanik Dalam
Data 1	17	91	6
Data 2	6	4	9
Data 3	63	41	7
Data 4	9	36	21
Data 5	6	32	1
Data 6	34	12	42
Data 7	23	14	8
Data 8	33	1	16
Data 9	76	23	1
Data 10	18	80	21
Data 11	17	32	13
Data 12	17	77	14
Data 13	1	56	31
Data 14	19	5	42
Data 15	18	18	6
Data 16	55	32	67
Data 17	122	4	8
Data 18	65	61	8
Data 19	18	6	20
Data 20	12	81	27

D. Evaluasi Aplikasi

Evaluasi aplikasi dilakukan dengan menguji akurasi aplikasi dengan data uji. Aplikasi yang diuji memiliki jumlah data latih yang berbeda-beda. Hal ini dapat menguji kebenaran keputusan atau pengenalan pola yang dilakukan oleh aplikasi. Keputusan aplikasi akan dibandingkan dengan fakta kelas yang sesungguhnya (*benchmarking*) untuk mengukur keakuratan aplikasi.

Evaluasi berikutnya adalah percobaan dengan variasi data latih. Hal ini akan menunjukkan korelasi antara banyaknya data latih terhadap akurasi keputusan.

IV. HASIL DAN PEMBAHASAN

Eksperimen dilakukan untuk menguji kinerja aplikasi yang dibangun. Tabel 3 menyajikan hasil pengujian. Berdasarkan percobaan yang dilakukan sebanyak empat kali, dimana masing-masing percobaan memiliki jumlah data latih yang berbeda-beda diperoleh hasil yang dapat menggambarkan performa aplikasi.

TABEL 3. HASIL PENGUJIAN

Data	Percobaan				Fakta
	Uji 1	Uji 2	Uji 3	Uji 4	
Data 1	siaga	Siaga	siaga	siaga	siaga
Data 2	waspada	Waspada	normal	normal	normal
Data 3	waspada	Waspada	waspada	waspada	waspada
Data 4	normal	Normal	normal	normal	normal
Data 5	normal	Normal	waspada	normal	normal
Data 6	waspada	Waspada	waspada	waspada	waspada

Data 7	waspada	Waspada	waspada	waspada	siaga
Data 8	waspada	Waspada	waspada	waspada	waspada
Data 9	waspada	Waspada	waspada	waspada	waspada
Data 10	siaga	Siaga	siaga	siaga	siaga
Data 11	normal	Normal	normal	normal	normal
Data 12	siaga	Waspada	siaga	siaga	siaga
Data 13	normal	Normal	normal	normal	normal
Data 14	normal	Waspada	waspada	siaga	siaga
Data 15	normal	Waspada	waspada	waspada	waspada
Data 16	siaga	waspada	waspada	siaga	waspada
Data 17	waspada	waspada	waspada	waspada	waspada
Data 18	waspada	waspada	waspada	waspada	waspada
Data 19	normal	waspada	normal	waspada	waspada
Data 20	siaga	siaga	normal	normal	normal

Pada percobaan pertama yang menggunakan 15 data-set, dari 20 percobaan, 13 data uji sesuai dengan fakta yang ada, dan 7 lainnya masih salah. Dengan demikian tingkat akurasi adalah 65%. Percobaan ke-2 menggunakan 20 data-set. Hasilnya adalah 15 data dari 20 data yang diuji memberikan hasil yang sama dengan fakta sebenarnya, dengan 5 hasil lainnya tidak sesuai dengan fakta. Tingkat akurasi sebesar 75%. Percobaan ke-3 menggunakan jumlah data-set sebanyak 25 data. Hasil yang sesuai fakta sebanyak 16 data, dan 4 yang lainnya masih tidak sesuai dengan fakta. Tingkat akurasi sebesar 80%. Percobaan ke-4 menggunakan data-set sebanyak 30 data. Hasilnya 18 data dari 20 data tersebut sesuai dengan fakta, sementara 2 lainnya masih salah klasifikasi. Dengan demikian tingkat akurasi mencapai 90%.

Grafik pada Gambar 2 menyimpulkan hubungan antara jumlah data latih dengan tingkat akurasi. Semakin besar jumlah data latih akan meningkatkan akurasi pengenalan oleh aplikasi.

Gambar 2. Grafik Hasil Pengujian

V. KESIMPULAN

Penelitian ini telah menerapkan *Naïve Bayesian Classifier* sebagai *Early Warning System* status gunung berapi berdasarkan fitur nilai Tektonik Jauh, Vulkanik Dangkal, dan Vulkanik Dalam suatu gunung berapi, kemudian dapat dikenali status suatu gunung berapi apakah termasuk salah satu kelas yaitu siaga, normal, atau waspada. Tingkat akurasi tertinggi mencapai 90% dengan jumlah data latih sebanyak 30 data. Penambahan data latih memberi dampak yang signifikan terhadap akurasi aplikasi. Selanjutnya dapat dikembangkan *real time system* untuk memantau status gunung berapi.

REFERENSI

- [1] Lessy, M. Zonasi Gempa Bumi di Indonesia, http://www.academia.edu/4517794/Zonasi_Gempa_bumi_di_Indonesia diakses, 2 November 2013.
- [2] Butaru, R. Posisi Indonesia dan Kerentanan terhadap Bencana. http://bulletin.penataanruang.net/upload/data_artikel/posisi%20indonesia.pdf, diakses 1 November 2013.
- [3] Duda, O. Richard, Hart, P. E, and Stork, D. G, "Pattern Classification (2nd ed)", John Wiley & Sons, 2000
- [4] Theodoridis, Sergios, Koutroumbas, Konstantin, Pattern Recognition (4th ed), Prentice Hall, 2005
- [5] Zeng T; Zheng L; Yang Li; Chen X; Long T. Offline Performance Prediction of PDAF With Bayesian Detection for Tracking in Clutter. IEEE Transactions on Signal Processing. Year: 2013, Volume: 61, Issue: 3 Pages: 770 - 781
- [6] Graham, Paul, 2003. Better Bayesian filtering. Downloaded at <http://vwww.paulgraham.com/better.html> Salib, Michael. 2002. "Meat Slicer: Spam Classification with Naive Bayes and Smart Heuristics".
- [7] Lingyan D; Yuqiao Wen; Jiang Ma. Dual tree complex wavelet transform and Bayesian estimation based denoising of poisson-corrupted X-ray images. 2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP). Year: 2013 Pages: 598 - 603
- [8] proxy.vsi.esdm.go.id/index.php, diakses tanggal 5 november 2013