ISSN cetak 2087-1716
ISSN online 2548-7779

ILKOM Jurnal Ilmiah Volume 9 Nomor 2 Agustus 2017

ALGORITMA C4.5 UNTUK PEMODELAN DAERAH RAWAN BANJIR STUDI KASUS KABUPATEN KARAWANG JAWA BARAT

Ahmad Khusaeri[1], Septian Ilham[2], Desi Nurhasanah[3], Derrenz Delpidat[4] , Anggri[5], Aji Primajaya[6], Betha Nurina Sari[7]

[1]ahmad.khusaeri15012@student.unsika.ac.id , [2]septian.ilham15134@student.unsika.ac.id , [3]desi.nurhasanah15033@student.unsika.ac.id, [4]derrenz.delpidat15032@student.unsika.ac.id, [5]anggril@gmail.com, [6] aji.primajaya@staff.unsika.ac.id, [7]betha.nurina@staff.unsika.ac.id

[1,2,3,4,5,6,7]Universitas Singaperbangsa Karawang

Abstrak

Karawang merupakan kota yang dijuluki dengan kota lumbung padi dan industri. Kemajuan kota Karawang ditentukan dengan situasi dan kondisi kota yang terkendali. Banjir merupakan salah satu hal yang dapat memperhambat kemajuan kota Karawang. Dengan membuat model daerah rawan banjir, dapat mempermudah pemerintah dalam memprediksi daerah – daerah yang rawan banjir berdasarkan data – data yang telah diolah. Algoritma C4.5 merupakan salah satu metode dalam mengklasifikasikan dataset. Pembagian dataset dilakukan dengan menggunakan metode 10-*fold cross validation* dimana setiap bagian (*fold*) akan dijadikan data *testing* sedangkan n-1 akan dijadikan sebagai data training. Data yang digunakan dalam pembuatan model adalah data banjir, kepadatan penduduk, ketinggian wilayah, jarak pusat daerah dengan sungai dan rata – rata curah hujan. Tingkat akurasi yang dihasilkan dengan menggunakan algoritma C4.5 adalah 84.385%, sedangkan *Recall* dan *Precision* yang dihasilkan sebesar 0.95 *dan* 0.87. pohon keputusan yang dihasilkan dapat dijadikan sebagai aturan dalam pembuatan visualisasi WEB GIS.

Kata kunci: GIS, Algoritma C4.5, *K-Fold Cross Validation*, Banjir, Karawang

1. Pendahuluan

Karawang merupakan salah satu kabupaten yang berada di provinsi Jawa Barat. Letak geografis Kabupaten Karawang terletak antara 107°02' – 107°40' BT dan 5° 56'- 6° 34'LS, termasuk daerah dataran yang relatif rendah, mempunyai variasi ketinggian wilayah antara 0-1.279 meter di atas permukaan laut dengan kemiringan wilayah 0-2°, 2-15°,15-40° dan di atas 40° dengan suhu rata – rata 27° C. ketinggian yang relatif rendah (25 m dpl) terletak pada Karawang bagian utara. Sedangkan untuk Karawang bagian selatan memiliki ketinggian antara 26 – 1.200 dpl. Selain letak geografisnya, Kabupaten Karawang pun memiliki luas wilayah kurang lebih 1.753,27 Km^2 [1].

Menurut data yang ada pada kantor BPS Karawang, hampir setiap tahun curah hujan yang ada di Kabupaten Karawang tinggi dan menyebabkan banjir. Banjir merupakan salah satu aspek yang dikhawatirkan oleh masyarakat dan pemerintah Kabupaten Karawang. Kekhawatiran masyarakat dan pemerintah Kabupaten Karawang akan terjadinya banjir didasari dari letak geografis Kabupaten Karawang yang memiliki ketinggian permukaan tanah yang memasuki kategori daerah rawan banjir karena Kabupaten Karawang merupakan daerah dengan dataran rendah dibanding dengan daerah lain di Pulau Jawa.

Kerugian yang ditimbulkan oleh banjir, terutama kerugian tidak langsung, mungkin menempati urutan pertama atau kedua setelah gempa bumi atau tsunami [2]. Dengan adanya banjir membuat sektor industri dan pertanian yang sangat berperan penting dalam memajukan kualitas ekonomi masyarakat Kabupaten Karawang menjadi menurun. Mengingat sektor yang menjadi sektor penting dalam meningkatkan kualitas ekonomi masyarakat Kabupaten Karawang adalah di sektor industri dan sektor pertanian. Sektor industri akan terhambat dengan adanya banjir karena sektor industri bergantung kepadaa sarana prasarana yang tidak terganggu. Sedangkan sektor pertanian akan merasakan dampak dari terjadinya banjir di Kabupaten Karawang mengingat kelancaran sektor pertanian tergantung dari kondisi cuaca dan kondisi tempat.

Dalam upaya untuk mengantisipasi banjir di Kabupaten Karawang, perlu adanya kajian mengenai kerentanan banjir. Prediksi potensi banjir pada daerah Karawang perlu dilakukan sebagai bagian dari kajian kerentanan banjir. Pembuatan model prediksi rawan banjir merupakan salah satu alternatif yang dapat dilakukan dalam upaya melaksanakan pencegahan terhadap daerah rawan banjir. Pembuatan model daerah rawan banjir dapat dilakukan dengan menggunakan algoritma C4.5 dan akan menghasilkan sebuah pohon keputusan.

Copyright © 2017 – ILKOM Jurnal Ilmiah -- All rights reserved |

ISSN cetak 2087-1716
ISSN online 2548-7779

2. Metode

2.1. Data dan Area Studi

Data yang digunakan saat penelitian adalah data spasial. Data Spasial merupakan data yang menunjuk posisi geografi dimana setiap karakteristik memiliki satu lokasi yang harus ditentukan dengan cara yang unik [3]. Model data vektor menggunakan titik dan koordinat x, y untuk membangun fitur spasial titik, terdiri atas titik dan *edge* (garis) yang menghubungkan titik-titik ini ke bentuk *polygon*. Model data *raster* menggunakan sebuah *grid* untuk merepresentasikan variasi spasial dari sebuah fitur. Setiap sel dalam *grid* memiliki nilai yang berhubungan dengan karakteristik fitur spasial pada lokasi itu. Data *raster* sangat cocok digunakan untuk merepresentasikan fitur spasial yang kontinu seperti curah hujan dan ketinggian [4].

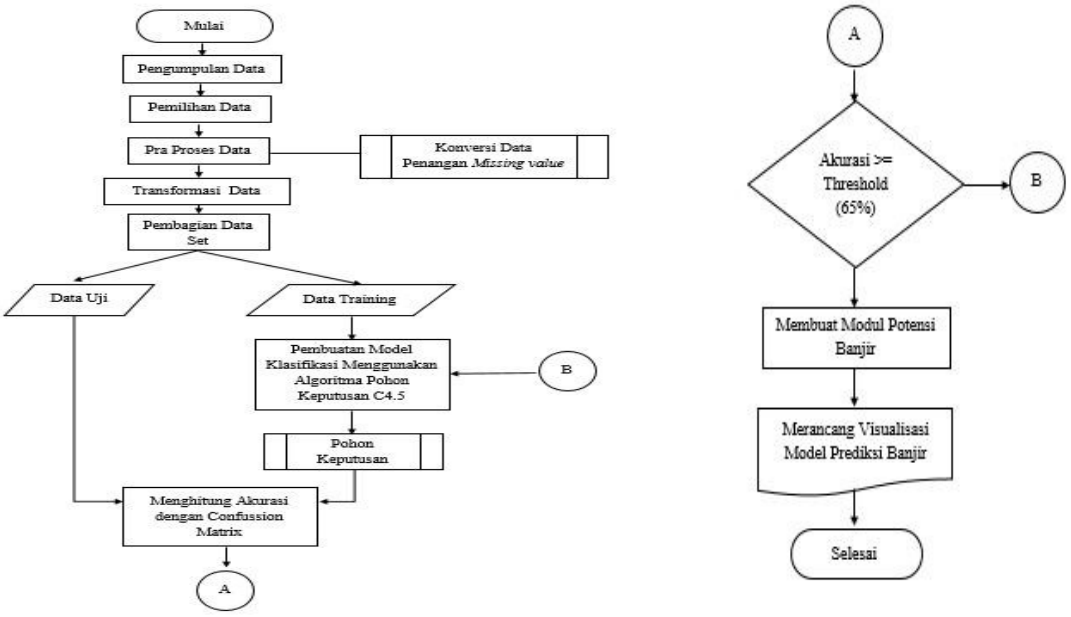
Area studi yang digunakan adalah Kabupaten Karawang. Ada tiga faktor yang sangat berpengaruh terhadap banjir yaitu elemen meteorologi, karakteristik DAS dan faktor manusia [5] yang selanjutnya dijadikan sebagai parameter banjir. Data parameter prediksi potensi banjir yaitu data curah hujan, data kepadatan penduduk serta data ketinggian permukaan tanah. Data curah hujan yaitu data frekuensi hujan yang terjadi setiap sebulan sekali yang terjadi di Kabupaten Karawang. Selain curah hujan, data kepadatan penduduk pun menjadi salah satu acuan dalam prediksi potensi banjir. Data kepadatan penduduk didapat dari Badan Statistika Kabupaten Karawang yang diperoleh dari setiap kecamatan di Kabupaten Karawang.

Tabel 1. Sumber Data

No.	Data	Sumber
1	Data Banjir tahun 2013	Badan Nasional Penanggulangan Bencana
2	Data Rata Rata Curah Hujan Tahun 2013	BPS Kabupaten Karawang
3	Data Jarak Centroid Desa dengan Sungai	Perhitungan dengan menggunakan Quantum GIS
4	Data Kepadatan Penduduk Tahun 2013	BPS Kabupaten Karawang
5	Data Ketinggian Wilayah Tahun 2013	BPS Kabupaten Karawang

2.2. Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang sistematis. Adapun tahapan penelitiannya sebagai berikut.



Gambar 1. Tahapan Penelitian

Copyright © 2017 – ILKOM Jurnal Ilmiah -- All rights reserved |

ISSN cetak 2087-1716
ISSN online 2548-7779

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Tahap pertama yaitu proses pengumpulan data dari setiap parameter yang dibutuhkan. Pengumpulan data harus memenuhi beberapa prinsip yaitu mengumpulkan data selengkap – lengkapnya, mempertimbangkan ketepatan data dan kebenaran data [6]. Pengumpulan data dilakukan dengan melaksanakan observasi pada BPS Karawang dan BPBD Kabupaten Karawang. Selain itu, dilakukan pengumpulan data pada setiap website resmi dari setiap instansi yang dapat menunjang penelitian. Data yang dikupulkan merupakan data tahun 2013 dari setiap parameter. Adapun data yang dikumpulkan sebagai berikut.

3.2 Pemilihan Data

Tahap kedua yaitu pemilihan data dengan atribut yang sesuai dengan kebutuhan penelitian. Pada tahap ini, terdapat berbagai macam atribut di dalam sebuah data, hanya beberapa atribut dari data yang akan digunakan sesuai dengan kebutuhan dari parameter. Setelah itu, dirapihkan sesuai setiap parameter data yang telah dipilih. Adapun atribut yang digunakan adalah data banjir, data rata – rata curah hujan, data kepadatan penduduk, data ketinggian wilayah per desa dan data jarak pusat desa dengan sungai. Untuk data banjir, data rata – rata curah hujan, data kepadatan penduduk dan data ketinggian wilayah per desa menggunakan data numerik sedangkan data banjir akan ditranformasi kedalam data kategorikal.

3.3 Pra Proses Data

Tahapan selanjutnya adalah praproses data. Di dalam praproses data dilakukan hal – hal sebagai berikut.

1. Mengonversi nilai data

   Nilai data jarak pusat desa ke sungai semula dalam bentuk satuan milimeter. Untuk mempermudah dalam proses data, nilai data tersebut dikonversi menjadi meter dengan cara mengalikan setiap nilai dengan pengali 50000. Selain itu, nilai data kepadatan penduduk semula dalam bentuk satuan hektare dikonversikan menjadi kilometer persegi dengan membagi nilai data dengan pembagi 100. Adapun contoh data yang belum dikonversikan dan yang sudah dikonversikan sebagai berikut

2. Penanganan *missing value.*

   Data yang mengalami *missing value* adalah data rata – rata curah hujan per kecamatan. Untuk mengatasi *missing value* dilakukan dengan cara memasukkan nilai rata - rata sesuai desanya masing - masing. *Missing value* harus diisi karena dapat mempengaruhi terhadap hasil dari pohon keputusan. Adapun contoh penanganan *missing value* dengan metode rata-rata sebagai berikut.

3.4 Transformasi Data

Pada tahap ini dilakukan perubahan bentuk data banjir dari informasi menjadi data kategorikal. Hal ini dengan mengategorikan desa yang pernah terjadi banjir dan desa yang belum pernah terkena banjir. Untuk daerah yang pernah terkena banjir masuk ke dalam kategori “Ya” sedangkan desa yang belum pernah terkena banjir masuk ke dalam kategori “Tidak”.

3.5 Pembagian Data Set

Pembagian data set dilakukan dengan metode *10-fold cross validation* menggunakan software Weka 3.8.1. Data Set dibagi menjadi data testing dan data training. Setiap bagian (*fold*) dijadikan sebagai data uji dan *n*-1 *fold* dijadikan data latih.

3.6 Pembuatan Model Pohon Keputusan

Pada tahap keenam dilakukan pembuatan model pohon keputusan. Pembuatan pohon keputusan dilakukan dengan algoritma C4.5 yang diimplementasikan pada Weka 3.8.1 dengan modul J48. Untuk mendapatkan pohon keputusan terlebih dahulu data set diklasifikasikan dengan *10-fold cross validation.* Data yang digunakan adalah data numerik dengan class banjir dengan bentuk kategorikal ya atau tidak. Algoritma C4.5 menghasilkan 12 aturan. Adapun contoh aturan tersebut adalah sebagai berikut.

Copyright © 2017 – ILKOM Jurnal Ilmiah -- All rights reserved |

ISSN cetak 2087-1716
ISSN online 2548-7779

Tabel 2. Hasil Dari Pohon Keputusan

No	Aturan
1	Jika Rata – Rata Curah Hujan Lebih dari 23..2424 maka tidak berpotensi banjir.
2	Jika Rata – Rata Curah Hujan Kurang Dari 23.2424 dan Ketinggian lebih dari 24 maka tidak berpotensi banjir
3	Jika Rata – Rata Curah Hujan Kurang Dari 23.2424, Ketinggian Kurang Dari Sama Dengan 24, Rata – Rata Curah Hujan Lebih dari 20.39 dan Ketinggian Kurang Dari Sama Dengan 13.2 Maka Berpotensi Banjir.

Beberapa aturan diatas diperoleh dari pohon keputusan yang dihasilkan melalui software weka 3.8.1.

3.7 Perhitungan Akurasi Pohon Keputusan

Tahap ketujuh adalah menghitung akurasi pohon keputusan. Dengan menggunakan pengujian *10-fold cross validation* menghasilkan *confussion matrix* sebagai berikut.

Tabel 3. *Confussion Matrix*

		Kelas Hasil Prediksi		Total
		Tidak	Ya	
Kelas Aktual	Tidak	243	11	254
	Ya	36	11	47
Total		279	22	301

Adapun cara untuk menghitung akurasi pohon keputusan adalah sebagai berikut.

$$\textbf{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Keterangan :

TP = *True Positive* (Kelas Aktual Ya dan Hasil Prediksi Ya)
TN = *True Negative* (Kelas Aktual Tidak dan Hasil Prediksi Tidak)
FP = *False Positive* (Kelas Aktual Ya dan Hasil Prediksi Tidak)
FN = *False Negative* (Kelas Aktual Tidak dan Hasil Prediksi Ya)

$$\textbf{Akurasi} = \frac{11 + 243}{11 + 243 + 36 + 11}$$

$$\textbf{Akurasi} = \frac{254}{301}$$

Akurasi = 0,84385 x 100

Akurasi = 84, 385 %

Jadi akurasi pohon keputusan yang menunjukkan keakuratan pohon keputusan dalam mengklasifikasikan daerah rawan banjir adalah 84, 385 %. Tingkat akurasi ini sudah melewati threshold 65 % , itu artinya pohon keputusan ini dapat dijadikan sebagai aturan dalam membuat modul klasifikasi daerah rawan banjir.

Copyright © 2017 – ILKOM Jurnal Ilmiah -- All rights reserved

ISSN cetak 2087-1716
ISSN online 2548-7779

4. Kesimpulan dan saran

4.1. Kesimpulan

Dari uraian yang telah dijelaskan diatas, dapat ditarik beberapa kesimpulan yang didapat sebagai berikut.

1. Tingkat akurasi yang dihasilkan dengan menggunakan algoritma C4.5 adalah sebesar 84.385%.
2. *Precision* dan *Recall* dari pohon keputusan tersebut adalah 0.871 dan 0.957.

4.2. Saran

1. Untuk meningkatkan tingkat akurasi dari pohon keputusan, data yang digunakan dapat lebih banyak lagi dan ditambahkan dengan beberapa atribut yang berbeda.
2. Dapat dilakukan penggabungan beberapa algoritma untuk menentukan tingkat akurasi yang lebih akurat dan tinggi.

Daftar Pustaka

[1] KARAWANG, B. K. (2014). *Karawang Dalam Angka 2014.* Karawang: Bps Kabupaten Karawang.
[2] Rosydie, A. (2013). Banjir : Fakta Dan Dampaknya, Serta Pengaruh Dari Perubahan Guna Lahan. *Perencanaan Wilayah Dan Kota*, 246.
[3] Dewi Handayani U.N, R. d. (2005). Pemanfaatan Analisis Spasial untuk Pengolahan Data Spasial. *Jurnal Teknologi Informasi DINAMIK*, 109.
[4] Amri, K. (2014). Sistem Informasi Geografis Berbasis Web Untuk Klasifikasi Kemunculan Titik Api Di Provinsi Riau Menggunakan Pohon Keputusan C4.5.
[5] Hamdani, H., Permana, S. (2014). Analisa Daerah Rawan Banjir Menggunakan Aplikasi Sistem Informasi Geografis. *Jurnal Kontruksi*, 2.
[6] Ig. Dodiet Aditya S, S. (2013). *DATA dan Metode Pengumpulan Data Penelitian.* Surakarta: Poltekes Surakarta.

Copyright © 2017 – ILKOM Jurnal Ilmiah -- All rights reserved |