

Analisa Datamining dengan Metode Klasifikasi C4.5 Sebagai Faktor Penyebab Tanah Longsor

Afrialita Widiastari^{1,*}, Solikhun², Irawan¹

¹ STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

² AMIK Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

Email: ^{1,*}afrialitawidiastari56@gmail.com, ²solikhun@amiktunasbangsa.ac.id, ³irawaniwan56@gmail.com

Submitted: 10/05/2021; Accepted: 28/05/2021; Published: 28/05/2021

Abstrak—Tanah Longsor adalah peristiwa geologi dimana terjadi pergerakan tanah seperti jatuhnya bebatuan atau gumpalan besar pada tanah. Longsor kerap terjadi saat hujan, meski tak selalu. Selain itu, longsor umumnya terjadi di area dengan lereng curam. Dengan metode C4.5 dapat digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial, hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru. Data yang diperoleh dari pihak Badan Nasional Penanggulangan Bencana mengenai faktor penyebab tanah longsor tersebut dapat menghasilkan nilai accuracy sebesar 77,78% artinya aturan atau rule yang dihasilkan mendekati 100% dapat disimpulkan bahwa untuk mengklasifikasi faktor penyebab tanah longsor yang menggunakan C4.5 dengan melihat node tertinggi yaitu kelerengan Curam. Untuk mengetahui faktor tersebut dapat memberikan masukan kepada Instansi Badan Penanggulangan Bencana agar lebih peduli terhadap faktor yang menyebabkan tanah longsor tersebut.

Kata Kunci: Data Mining; C4.5; Faktor Penyebab Tanah Longsor

Abstract—Landslides are geological events where soil movement occurs such as falling rocks or large lumps on the ground. Landslides often occur when it rains, although not always. In addition, landslides generally occur in areas with steep slopes. With the C4.5 method it can be used to classify data that has numeric and categorical attributes. The results of the classification process in the form of rules can be used to predict the value of the discrete type attribute from a new record. Data obtained from the National Disaster Management Agency regarding the factors causing landslides can produce an accuracy value of 77.78%, meaning that the resulting rules or rules are close to 100%, it can be concluded that to classify the factors causing landslides using C4.5 by looking at the node the highest is slope. To find out these factors can provide input to the Disaster Management Agency to be more concerned with the factors that caused the landslides.

Keywords: Data Mining; C4.5; Factors Causing Landslides

1. PENDAHULUAN

Bencana merupakan suatu peristiwa yang mengancam dan mengganggu kehidupan dan penghidupan masyarakat yang mengakibatkan timbulnya korban jiwa manusia, kerusakan lingkungan, kerugian harta benda, dan dampak psikologis yang disebabkan, baik oleh faktor alam dan faktor non alam maupun faktor manusia. Bencana tidak terjadi begitu saja namun ada faktor kesalahan dan kelalaian manusia dalam mengantisipasi alam dan kemungkinan bencana yang dapat menyimpannya. “Tanah longsor merupakan bencana yang terjadi akibat intensitas hujan yang tinggi yang terjadi lebih dari satu hari”. [1]. Struktur tanah yang labil dan tidak merata sangat rawan terjadinya tanah longsor. Masyarakat yang tinggal di lereng gunung curam, menghadapi resiko kemungkinan terjadinya tanah longsor. Tanah longsor juga dapat dipicu oleh getaran gempa hingga merontokkan struktur tanah di atasnya.

Bencana tanah longsor yang pernah terjadi di Indonesia seperti longsor di Sungai Bohorok, Sumatera Utara pada bulan November 2003 dikarenakan curah hujan yang tinggi. Dengan kejadian tersebut maka perlu dilakukan klasifikasi faktor penyebab bencana tanah longsor di Indonesia dengan mengumpulkan data-data yang terkait berdasarkan penyebab bencana tanah longsor tersebut. Data-data tersebut diperoleh dari Data Informasi Bencana Indonesia (DIBI) BNPB yang terdiri dari beberapa variabel terkait faktor yang menimbulkan bencana tanah longsor yang terdiri dari Kelerengan, Jenis Tanah, Curah Hujan, Topografi dan Longsoran. Dengan meningkatnya bencana tanah longsor tersebut, maka dapat ditentukan penyebab bencana tanah longsor. Dalam mengklasifikasikan faktor penyebab tanah longsor tersebut digunakan metode *datamining* dengan algoritma C4.5.

Datamining merupakan proses yang menggunakan teknik statistik, matematika, *artificial intelegent* dan *machine learning* untuk menginteraksi, mengidentifikasi informasi bermanfaat yang terikat dari berbagai database besar” [1]. Algoritma C4.5 adalah “algoritma klasifikasi data bertipe pohon keputusan” [2]. Pohon keputusan Algoritma C4.5 dibangun dengan tahapan yang meliputi pemilihan atribut sebagai node awal, membentuk cabang untuk setiap nilai dan mempartisi kasus dalam cabang. Tahapan-tahapan ini akan diulangi untuk setiap cabang sampai semua kasus pada cabang habis terpartisi. Dari penyelesaian tersebut maka akan didapatkan beberapa rule [3], [4].

Dari penelitian ini diharapkan dapat membantu pihak BNPB (Badan Nasional Penanggulangan Bencana) dalam memberi solusi terbesar yang menyebabkan bencana tanah longsor.

2. METODOLOGI PENELITIAN

2.1 Data Mining

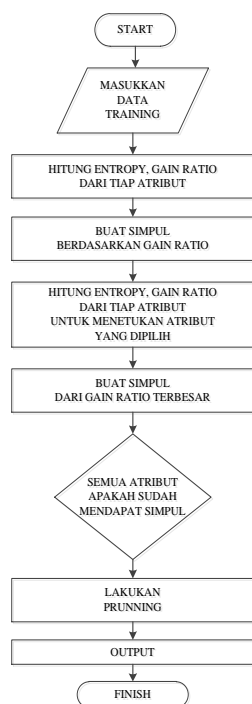
Data Mining Merupakan suatu rangkaian proses yang dapat dibagi menjadi beberapa tahap yang mana tahap-tahap tersebut bersifat *interaktif* di mana pemakai terlibat langsung atau dengan perantaraan *knowledge base* [5]. *Data Mining* sering juga disebut Knowledge discovery in database (KDD) merupakan meliputi kegiatan pengumpulan, pemakaian data historis menemukan pola atau hubungan dalam set data berukuran besar. “Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan” [6], [7].

2.2 Rapidminer

Rapidminer adalah platform perangkat ilmu data yang dikembangkan oleh perusahaan ternama yang menyediakan lingkungan terintegrasi untuk persiapan data pembelajaran mesin pembelajaran dalam pengembangan teks dan analisis prediktif. Memudahkan pengguna dalam menggunakan perangkat ini. Berbagai metode yang disediakan software rapidminer antara lain klasifikasi, klustering dan assosiasi [8].

2.3 Pemodelan Metode

Setiap metode atau algoritma memiliki pemodelannya masing-masing sesuai pengerjaan, pada penelitian ini penulis menggunakan metode C4.5 untuk mengklasifikasi faktor penyebab terjadinya tanah longsor.



Gambar 1. Flowchart metode C4.5.

3. HASIL DAN PEMBAHASAN

Permasalahan yang sudah dianalisa, selanjutnya penulis menentukan faktor penyebab bencana tanah longsor di Sumatera Utara dengan menggunakan *Data Mining* algoritma C4.5. Algoritma C4.5 dilakukan dengan merancang pohon keputusan (*decision tree*) dengan memilih atribut pada variabel data sebagai akar (*node*), membuat cabang untuk setiap nilai, membagi kasus dalam cabang, dan mengulangi proses sampai semua cabang memiliki kelas yang sama. Perhitungan *Entropy* dan *Gain* dilakukan untuk membentuk suatu keputusan. Membuat pohon keputusan terlebih dahulu menghitung jumlah kasus untuk keputusan Paham dan jumlah kasus keputusan Tidak paham. Berikut ini langkah-langkah dalam membentuk pohon keputusan dengan menggunakan Algoritma C4.5:

- Menghitung *Entropy* dari total data. Dari data penelitian diketahui jumlah kasus ada 36 *record*, diketahui status Besar yang paham ada 25 *record* dan status Kecil ada 11 *record* sehingga *Entropy* nilai akar total yang didapat:

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-25/36 * \log_2 (25/36)) + (-11/36 * \log_2 (11/36)) \\
 &= 0,8880
 \end{aligned}$$

- b) Menghitung nilai *Entropy* pada atribut Faktor Kelerengan dengan kategori Rendah, jumlah kasus terdapat 11 *record*, terdapat status Besar sebanyak 2 *record* dan status Kecil sebanyak 9 *record*.

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-2/11 * \log_2 (2/11)) + (-9/11 * \log_2 (9/11)) \\ &= 0,6840 \end{aligned}$$

- c) Seterusnya perhitungan *Entropy* pada Faktor Kelerengan sampai dengan kategori Curam. Selanjutnya menghitung nilai *Gain* pada Atribut Faktor Kelerengan dengan kategori Rendah terdapat 11 *record*, kategori Terjal terdapat 12 *record*, dan kategori Curam terdapat 13 *record* :

$$\begin{aligned} Gain(S,A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= 0,8880 - (((11/36)*0,6840) + ((12/36)*0,6500) + \\ &\quad ((13/36)*0)) \\ &= 0,4623 \end{aligned}$$

- d) Selanjutnya Menghitung *SplitInfo* pada atribut Faktor Kelerengan berdasarkan jumlah kasus pada Faktor Kelerengan.

$$\begin{aligned} SplitInfo &= \sum_{i=1}^k p(v_i | s) \log_2 p(v_i | s) \\ &= -((11/36 * \log_2 (11/36)) + (12/36 * \log_2 (12/36) + (13/36 * \log_2 \\ &\quad (13/36))) \\ &= 1,5816 \end{aligned}$$

- e) Kemudian menghitung Nilai *GainRatio* Faktor Kelerengan sebagai nilai yang digunakan sebagai pemecahan *Node* pada pohon keputusan.

$$\begin{aligned} RasioGain(s,j) &= \frac{Gain(s,j)}{SplitInfo(s,j)} \\ &= 0,4623 / 1,5816 \\ &= 0,2923 \end{aligned}$$

- f) Menghitung nilai *Entropy* pada atribut Faktor Jenis Tanah dengan kategori Aluvial, jumlah kasus terdapat 12 *record*, terdapat status Besar sebanyak 6 *record* dan status Kecil sebanyak 6 *record*.

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-6/12 * \log_2 (6/12)) + (-6/12 * \log_2 (6/12)) \\ &= 1,0000 \end{aligned}$$

- g) Seterusnya perhitungan *Entropy* pada Faktor Jenis Tanah sampai dengan kategori Podsol. Selanjutnya menghitung nilai *Gain* pada Atribut Faktor Jenis Tanah dengan kategori Aluvial terdapat 12 *record*, kategori Latosol terdapat 1 *record*, kategori Grumasol terdapat 2 *record*, kategori Mediteran terdapat 15 *record* dan kategori Podsol terdapat 6 *record* :

$$\begin{aligned} Gain(S,A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= 0,8880 - (((12/36)*1,0000) + ((1/36)*0) + \\ &\quad ((2/36)*0) + ((15/36)*0,5665) + ((6/36)*0)) \\ &= 0,3186 \end{aligned}$$

- h) Selanjutnya Menghitung *SplitInfo* pada atribut Faktor Jenis Tanah berdasarkan jumlah kasus pada Faktor Jenis Tanah

$$\begin{aligned} SplitInfo &= \sum_{i=1}^k p(v_i | s) \log_2 p(v_i | s) \\ &= -((12/36 * \log_2 (12/36)) + (1/36 * \log_2 (1/36)) + \\ &\quad (2/36 * \log_2 (2/36)) + (15/36 * \log_2 (15/36)) \\ &\quad + (6/36 * \log_2 (6/36))) \\ &= 1,8607 \end{aligned}$$

- i) Kemudian menghitung Nilai *GainRatio* Faktor Jenis Tanah sebagai nilai yang digunakan sebagai pemecahan *Node* pada pohon keputusan.

$$\begin{aligned} RasioGain(s,j) &= \frac{Gain(s,j)}{SplitInfo(s,j)} \\ &= 0,3186 / 1,8607 \\ &= 0,1712 \end{aligned}$$

- j) Menghitung nilai *Entropy* pada atribut Curah Hujan dengan kategori Rendah, jumlah kasus terdapat 3 *record*, terdapat status Besar sebanyak 1 *record* dan status Kecil sebanyak 2 *record*.

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-1/3 * \log_2 (1/3)) + (-2/3 * \log_2 (2/3)) \\ &= 0,9183 \end{aligned}$$

- k) Seterusnya perhitungan *Entropy* pada Faktor Curah Hujan sampai dengan kategori Sangat Tinggi. Selanjutnya menghitung nilai *Gain* pada Atribut Faktor Curah Hujan dengan kategori Rendah terdapat 3 *record*, kategori Sedang terdapat 2 *record*, kategori Tinggi terdapat 21 *record*, dan kategori Sangat Tinggi terdapat 10 *record* :

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

$$= 0,8880 - (((3/36)*0,9183) + ((2/36)*0) + ((21/36)*0,9183) + ((10/36)*0,7219)) = 0,1354$$

- l) Selanjutnya Menghitung *SplitInfo* pada atribut Faktor Curah Hujan berdasarkan jumlah kasus pada Faktor Curah Hujan.

$$\begin{aligned} SplitInfo &= \sum_{i=1}^k p(v_i | s) \log_2 p(v_i | s) \\ &= -((3/36.\log_2(3/36)) + (2/36.\log_2(2/36) + (21/36.\log_2(21/36) + (10/36.\log_2(10/36))) \\ &= 1,4973 \end{aligned}$$

- m) Kemudian menghitung Nilai *GainRatio* Faktor Curah Hujan sebagai nilai yang digunakan sebagai pemecahan *Node* pada pohon keputusan.

$$\begin{aligned} RasioGain(s,j) &= \frac{Gain(s,j)}{SplitInfo(s,j)} \\ &= 0,0752 / 1,4973 \\ &= 0,0503 \end{aligned}$$

- n) Menghitung nilai *Entropy* pada atribut Faktor Topografi dengan kategori Bukit Gelombang, jumlah kasus terdapat 11 *record*, terdapat status Besar sebanyak 4 *record* dan status Kecil sebanyak 7 *record*.

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i \cdot \log_2 p_i \\ &= (-4/11 \cdot \log_2 (4/11)) + (-7/11 \cdot \log_2 (7/11)) \\ &= 0,9457 \end{aligned}$$

- o) Seterusnya perhitungan *Entropy* pada Faktor Topografi sampai dengan kategori Tebing Tegak. Selanjutnya menghitung nilai *Gain* pada Atribut Faktor Topografi dengan kategori Bukit Gelombang terdapat 11 *record*, kategori Bukit Terjal terdapat 15 *record*, dan kategori Tebing Tegak terdapat 10 *record* :

$$\begin{aligned} Gain(S,A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= 0,8880 - (((11/36)*0,9457) + ((15/36)*0,7219) + ((10/36)*0,4690)) \\ &= 0,1679 \end{aligned}$$

- p) Selanjutnya Menghitung *SplitInfo* pada atribut Faktor Curah Hujan berdasarkan jumlah kasus pada Faktor Curah Hujan.

$$\begin{aligned} SplitInfo &= \sum_{i=1}^k p(v_i | s) \log_2 p(v_i | s) \\ &= -((11/36.\log_2(11/36)) + (15/36.\log_2(11/36) + (10/36.\log_2(10/36))) \\ &= 1,5622 \end{aligned}$$

- q) Kemudian menghitung Nilai *GainRatio* Faktor Curah Hujan sebagai nilai yang digunakan sebagai pemecahan *Node* pada pohon keputusan.

$$\begin{aligned} RasioGain(s,j) &= \frac{Gain(s,j)}{SplitInfo(s,j)} \\ &= 0,1679 / 1,5622 \\ &= 0,1075 \end{aligned}$$

Hasil dari perhitungan nilai *Entropy*, nilai *Gain*, *SplitInfo* dan *GainRatio* untuk tiap atribut dapat dilihat pada tabel 1 dengan keterangan SB =Sangat Baik, B =Baik, C =Cukup, K =Kurang dan SK =Sangat Kurang sebagai berikut :

Tabel 1. Perhitungan *Gain Node 1*

No	Atribut	Jumlah Kasus	Besar	Kecil	Entropy	Gain	Split Info	Gain Ratio	
1	Total	36	25	11	0,8880	0,4623	1,5816	0,2923	
	Kelerengan								
	Rendah	11	2	9	0,6840				
	Terjal	12	10	2	0,6500				
2	Jenis Tanah	Curam	13	13	0	0,3186	1,8607	0,1712	
		Aluvial	12	6	6				1,0000
		Latosol	1	0	1				
		Grumasol	2	0	2				
		Mediteran	15	13	2				0,5665
		Podsol	6	6	0				
3	Curah Hujan					0,0752	1,4973	0,0503	
		Rendah	3	1	2				0,9183

No	Atribut	Jumlah Kasus	Besar	Kecil	Entropy	Gain	Split Info	Gain Ratio	
4	Topografi	Sedang	2	2	0	0,1679	1,5622	0,1075	
		Tinggi	21	14	7				0,9183
		Sangat Tinggi	10	8	2				0,7219
		Bukit	11	4	7				0,9457
		Gelombang	15	12	3				0,7219
		Bukit Terjal	10	9	1				0,4690
	Tebing								
	Tegak								

Dari tabel 1 diketahui nilai *GainRatio* terbesar adalah atribut Kelerengan dengan nilai 0,2923 kategori Rendah dengan nilai *Entropy* 0,6840, kategori Terjal dengan nilai 0,6500 dan dijadikan simpul akar untuk mengulangi perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya. Hasil perhitungan nilai *Entropy* dan nilai *Gain* untuk atribut Faktor Kelerengan untuk kategori Rendah sebagai Node 1.1 dan Kategori Terjal Untuk Node 1.2, Hasil Node 1.1 dapat dilihat pada tabel 2. berikut:

Tabel 2. Perhitungan *Gain Node 1.1*

No	Atribut	Jumlah Kasus	Besar	Kecil	Entropy	Gain	Split Info	Gain Ratio	
1	Kelerengan = Rendah Jenis Tanah	11	2	9	0,6840	0,1348	1,4911	0,0904	
2	Curah Hujan	Aluvial	7	2	5	0,8631	0,1348	0,4150	0,3248
		Latosol	1	0	1				
		Grumasol	2	0	2				
		Mediteran	1	0	1				
		Podsol	0	0	0				
3	Topografi	Rendah	2	0	2	0,0383	0,8454	0,0453	
		Sedang	0	0	0				
		Tinggi	7	2	5				0,8631
		Sangat Tinggi	2	0	2				
		Bukit Gelomban g	8	1	7				0,5436
		Bukit Terjal	3	1	2				0,9183
	Tebing	0	0	0					
	Tegak								

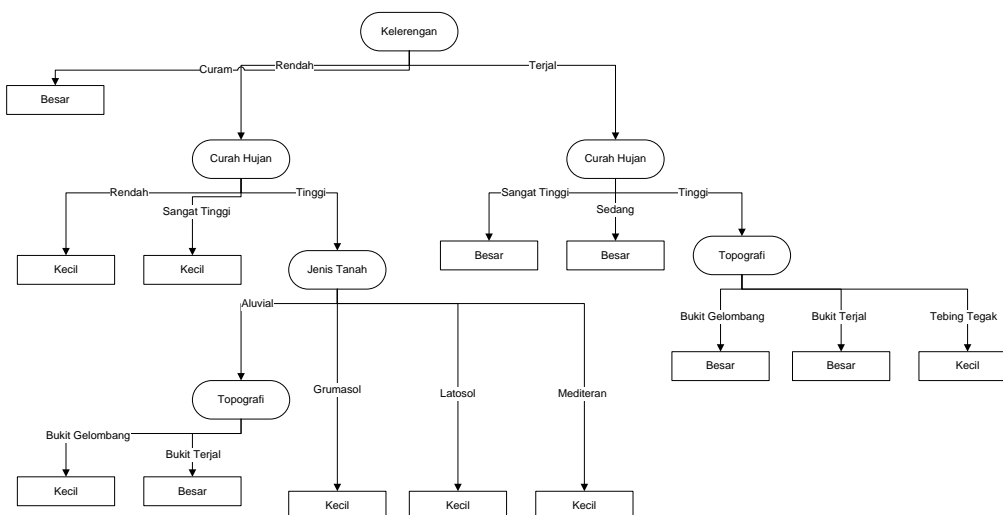
Dari tabel 2 diketahui nilai *GainRatio* terbesar adalah atribut Curah Hujan dengan nilai 0,3248 kategori Tinggi dengan nilai *Entropy* 0,8631, dan dijadikan simpul akar untuk mengulangi perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya. Hasil perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya untuk atribut Faktor Kelerengan Rendah dan Curah Hujan Tinggi sebagai Node 1.1.1 dapat dilihat pada tabel 3. berikut:

Tabel 3. Perhitungan *Gain Node 1.1.1*

No	Atribut	Jumlah Kasus	Besar	Kecil	Entropy	Gain	Split Info	Gain Ratio	
1	Kelerengan = Rendah, Curah Hujan = Tinggi Jenis Tanah	7	2	5	0,8631	0,4696	1,8424	0,2549	
		Aluvial	3	2	1				0,9183
		Latosol	1	0	1				

No	Atribut	Jumlah Kasus	Besar	Kecil	Entropy	Gain	Split Info	Gain Ratio	
2	Topografi	Grumasol	2	0	2	0,0617	0,8631	0,0715	
		Mediterran	1	0	1				
		Podsol	0	0	0				
		Bukit	5	1	4				0,7219
		Gelombang							
		Bukit Terjal	2	1	1				1,0000
		Tebing Tegak	0	0	0				

Dari tabel 3 diketahui nilai *GainRatio* terbesar adalah atribut Jenis Tanah dengan nilai 0,2549 kategori Aluvial dengan nilai *Entropy* 0,9183, dan dijadikan simpul akar untuk mengulangi perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya. Hasil perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya untuk atribut Faktor Kelerengan Rendah, Curah Hujan Tinggi dan Jenis Tanah Aluvial sebagai Node 1.1.1.1. selanjutnya mencari nilai *GainRatio* terbesar yang dapat dijadikan simpul akar untuk mengulangi perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya. Sehingga didapat bentuk pohon keputusan yang dapat dilihat pada Gambar 2 sebagai berikut :



Gambar 2. Pohon Keputusan

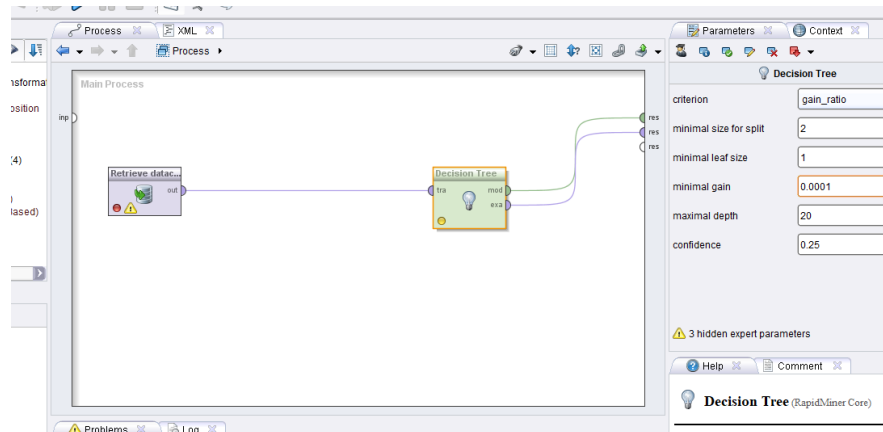
Berdasarkan bentuk Pohon Keputusan yang diperoleh pada gambar 1, aturan atau *Rule* model yang didapatkan adalah sebagai berikut:

1. Jika Kelerengan = Curam, maka hasilnya Besar {Besarnya = 13 dan Kecil = 0}
2. Jika Kelerengan = Rendah dan Curah Hujan = Rendah, maka hasilnya Kecil {Besarnya = 0 dan Kecil = 2}
3. Jika Kelerengan = Rendah dan Curah Hujan = Sangat Tinggi, maka hasilnya Kecil {Besarnya = 0 dan Kecil = 2}
4. Jika Kelerengan = Rendah, Curah Hujan = Tinggi, Jenis Tanah = Aluvial dan Topografi = Bukit Gelombang, maka hasilnya Kecil {Besarnya = 1 dan Kecil = 1}
5. Jika Kelerengan = Rendah, Curah Hujan = Tinggi, Jenis Tanah = Aluvial dan Topografi = Bukit Terjal, maka hasilnya Besar {Besarnya = 1 dan Kecil = 0}
6. Jika Kelerengan = Rendah, Curah Hujan = Tinggi dan Jenis Tanah = Grumasol, maka hasilnya Kecil {Besarnya = 0 dan Kecil = 2}
7. Jika Kelerengan = Rendah, Curah Hujan = Tinggi dan Jenis Tanah = Latosol, maka hasilnya Kecil {Besarnya = 0 dan Kecil = 1}
8. Jika Kelerengan = Rendah, Curah Hujan = Tinggi dan Jenis Tanah = Mediteran, maka hasilnya Kecil {Besarnya = 0 dan Kecil = 1}
9. Jika Kelerengan = Terjal dan Curah Hujan = Sangat Tinggi, maka hasilnya Besar {Besarnya = 4 dan Kecil = 0}
10. Jika Kelerengan = Terjal dan Curah Hujan = Sedang, maka hasilnya Besar {Besarnya = 2 dan Kecil = 0}
11. Jika Kelerengan = Terjal, Curah Hujan = Tinggi dan Topografi = Bukit Gelombang, maka hasilnya Besar {Besarnya = 2 dan Kecil = 0}
12. Jika Kelerengan = Terjal, Curah Hujan = Tinggi dan Topografi = Bukit Terjal, maka hasilnya Besar {Besarnya = 2 dan Kecil = 1}
13. Jika Kelerengan = Terjal, Curah Hujan = Tinggi dan Topografi = Tebing Tegak, maka hasilnya Kecil {Besarnya = 0 dan Kecil = 1}

3.1 Hasil Pengujian

Pada penelitian ini digunakan tools rapidminer dalam menguji data penelitian yang digunakan. Dalam proses rapidminer dilakukan *import* data menghasilkan informasi terdapat 36 *example* data yang merupakan jumlah sampel sebanyak 36 item, 1 *special attributes* yang merupakan 1 atribut special pada data yaitu atribut Bencana Longsor yang menggunakan jenis *label* dengan tipe *Binominal*, dan 4 *regular attributes* yang merupakan 4 atribut regular yang digunakan sebagai *node* akar pada pohon keputusan dengan jenis *attribute* dengan tipe *Polynomial*.

Langkah selanjutnya kembali ke tahapan desain pada *main process* untuk memproses data yang akan digunakan untuk mencari hasil keputusan. Langkah pertama *drag and drop* data yang telah di *importing* kedalam *main process* di menu *Repositories* pada *RapidMiner*. Langkah selanjutnya *drag and drop* operator *Decision Tree* dengan pengaturan pada operator tersebut kemudian hubungkan *port* pada seluruh operator seperti yang terlihat pada gambar 3 berikut:

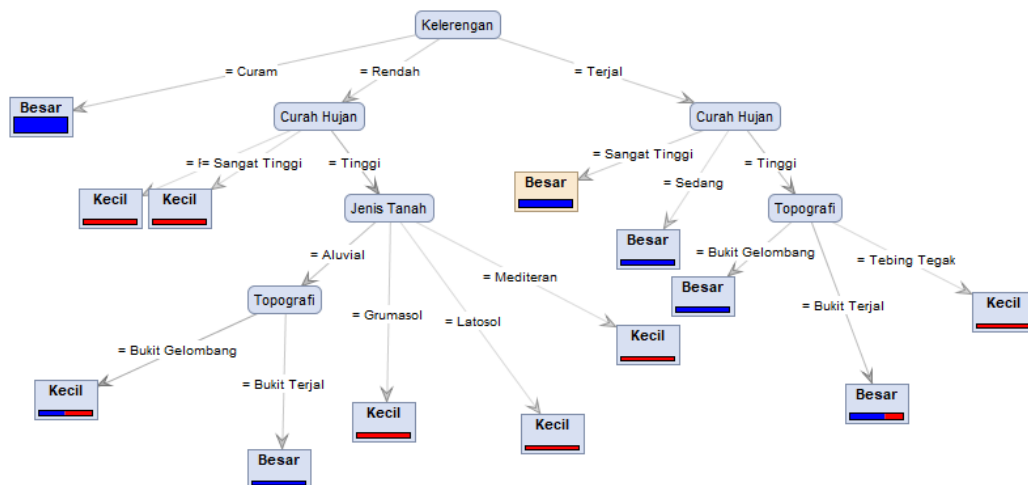


Gambar 3. Drag and Drop Operator Decision Tree

Gambar 3 dapat dijelaskan untuk operator *Decision Tree* terdiri dari beberapa pengaturan untuk Parameter. Penjelasan parameter pada operator *Decision Tree* dapat dilihat pada keterangan Gambar 2. Pengaturan parameter yang digunakan peneliti yaitu:

1. *Criterion* = *gain ratio*
2. *Minimal size for split* = 2
3. *Minimal leaf size* = 1
4. *Minimal Gain* = 0,0001
5. *Confidence* = 0,25

Klik tombol *Run* di *RapidMiner* maka muncul hasil model aturan dalam bentuk pohon keputusan yang dilakukan pada menu *Decision Tree* seperti yang terlihat pada gambar 4 berikut:



Gambar 4. Hasil Model Aturan Decision Tree di RapidMiner

Gambar 4 diatas menghasilkan aturan atau *rule* yang dilakukan di *RapidMiner* menu *Decision Tree* pada tab *Text View* seperti gambar 5 berikut:

```

Tree
Kelerengan = Curam: Besar {Besar=13, Kecil=0}
Kelerengan = Rendah
|  Curah Hujan = Rendah: Kecil {Besar=0, Kecil=2}
|  Curah Hujan = Sangat Tinggi: Kecil {Besar=0, Kecil=2}
|  Curah Hujan = Tinggi
|  |  Jenis Tanah = Aluvial
|  |  |  Topografi = Bukit Gelombang: Kecil {Besar=1, Kecil=1}
|  |  |  Topografi = Bukit Terjal: Besar {Besar=1, Kecil=0}
|  |  |  Jenis Tanah = Grumasol: Kecil {Besar=0, Kecil=2}
|  |  |  Jenis Tanah = Latosol: Kecil {Besar=0, Kecil=1}
|  |  |  Jenis Tanah = Mediteran: Kecil {Besar=0, Kecil=1}
Kelerengan = Terjal
|  Curah Hujan = Sangat Tinggi: Besar {Besar=4, Kecil=0}
|  Curah Hujan = Sedang: Besar {Besar=2, Kecil=0}
|  Curah Hujan = Tinggi
|  |  Topografi = Bukit Gelombang: Besar {Besar=2, Kecil=0}
|  |  Topografi = Bukit Terjal: Besar {Besar=2, Kecil=1}
|  |  Topografi = Tebing Tegak: Kecil {Besar=0, Kecil=1}
    
```

Gambar 5. Aturan atau Rule pada RapidMiner

Jika dilihat berdasarkan *text view* hasil pohon keputusan pada Gambar 4, peneliti membuat model aturan berupa teks berdasarkan Gambar 4 dengan penjelasan seperti berikut:

1. Jika Kelerengan = Curam, maka hasilnya Besar {Besar = 13 dan Kecil = 0}
2. Jika Kelerengan = Rendah dan Curah Hujan = Rendah, maka hasilnya Kecil {Besar = 0 dan Kecil = 2}
3. Jika Kelerengan = Rendah dan Curah Hujan = Sangat Tinggi, maka hasilnya Kecil {Besar = 0 dan Kecil = 2}
4. Jika Kelerengan = Rendah, Curah Hujan = Tinggi, Jenis Tanah = Aluvial dan Topografi = Bukit Gelombang, maka hasilnya Kecil {Besar = 1 dan Kecil = 1}
5. Jika Kelerengan = Rendah, Curah Hujan = Tinggi, Jenis Tanah = Aluvial dan Topografi = Bukit Terjal, maka hasilnya Besar {Besar = 1 dan Kecil = 0}
6. Jika Kelerengan = Rendah, Curah Hujan = Tinggi dan Jenis Tanah = Grumasol, maka hasilnya Kecil {Besar = 0 dan Kecil = 2}
7. Jika Kelerengan = Rendah, Curah Hujan = Tinggi dan Jenis Tanah = Latosol, maka hasilnya Kecil {Besar = 0 dan Kecil = 1}
8. Jika Kelerengan = Rendah, Curah Hujan = Tinggi dan Jenis Tanah = Mediteran, maka hasilnya Kecil {Besar = 0 dan Kecil = 1}
9. Jika Kelerengan = Terjal dan Curah Hujan = Sangat Tinggi, maka hasilnya Besar {Besar = 4 dan Kecil = 0}
10. Jika Kelerengan = Terjal dan Curah Hujan = Sedang, maka hasilnya Besar {Besar = 2 dan Kecil = 0}
11. Jika Kelerengan = Terjal, Curah Hujan = Tinggi dan Topografi = Bukit Gelombang, maka hasilnya Besar {Besar = 2 dan Kecil = 0}
12. Jika Kelerengan = Terjal, Curah Hujan = Tinggi dan Topografi = Bukit Terjal, maka hasilnya Besar {Besar = 2 dan Kecil = 1}
13. Jika Kelerengan = Terjal, Curah Hujan = Tinggi dan Topografi = Tebing Tegak, maka hasilnya Kecil {Besar = 0 dan Kecil = 1}

Berdasarkan aturan atau model dari *text view* Gambar 4 dapat dijelaskan bahwa faktor yang mempengaruhi pada *node* pertama adalah Faktor Kelerengan, *node* kedua Faktor Curah Hujan, *node* ketiga Faktor Jenis Tanah, *node* keempat Faktor Topografi. Untuk label Besar model aturan terbaik adalah jika Kelerengan Curam maka hasilnya Besar dengan jumlah atribut 13 dan label Kecil model aturan terbaik adalah jika Kelerengan Rendah, Curah Hujan Tinggi dan Jenis Tanah Latosol, maka hasilnya Kecil dengan jumlah atribut 1.

Untuk melihat keakuratan model aturan keterhubungan data faktor penyebab bencana longsor dengan menggunakan Algoritma C4.5 menggunakan Operator *X-Validation*. Operator *X-Validation* digunakan untuk memperkirakan seberapa akurat model yang akan dihasilkan dalam penelitian. Operator *X-Validation* memiliki dua *subprocess* yaitu *Training* yang digunakan untuk melatih sebuah model dan *Testing* yang digunakan untuk pengujian sebuah model. Akurasi Hasil pengujian model Algoritma C4.5 dapat dilihat pada Gambar 6 berikut:

Multiclass Classification Performance Annotations			
Table View Plot View			
accuracy: 77.50% +/- 16.27% (mikro: 77.78%)			
	true Besar	true Kecil	class precision
pred. Besar	22	5	81.48%
pred. Kecil	3	6	66.67%
class recall	88.00%	54.55%	

Gambar 6. Hasil Akurasi Model Algoritma C4.5

Gambar 6 menjelaskan nilai *Accuracy* sebesar 77,78.44%, artinya aturan atau *rule* yang dihasilkan mendekati 100% dengan *Class Precision* pada prediksi label Besar sebesar 81,49% dan prediksi label Kecil sebesar 66,57%.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan dapat disimpulkan bahwa untuk mengklasifikasikan faktor penyebab bencana tanah longsor menggunakan Algoritma C4.5 dengan melihat nilai Node tertinggi yang diperoleh dari nilai Rasio Gain (Gain Ratio) dengan hasil perolehan Node tertinggi yaitu Kelerengan. Hasil model aturan / model terbaik adalah jika Kelerengan Curam maka hasilnya Longsor Besar dengan hasil Label Besar bernilai 13 dan Label Kecil bernilai 0. Dapat disimpulkan pihak Pemerintah terkhusus untuk daerah di Sumatera Utara yang memiliki daerah Kelerengan Curam dapat berisiko terjadinya Bencana Longsor Besar.

REFERENCES

- [1] D. Setianingsih and R. F. Hakim, "Penerapan data mining dalam analisis kejadian tanah longsor di indonesia dengan menggunakan association rule algoritma apriori," *Pros. Semin. Nas. Mat. dan Pendidik. Mat. UMS 2015*, pp. 731–741, 2015.
- [2] A. Chandra, "Penerapan Data Mining Menggunakan Pohon Keputusan," vol. 6, no. 3, pp. 1–6, 2017.
- [3] E. Buulolo, N. Silalahi, Fadlina, and R. Rahim, "C4.5 Algorithm To Predict the Impact of the Earthquake," *Int. J. Eng. Res. Technol.*, vol. 6, no. 2, 2017.
- [4] A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Bagian 1 : Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)," *J. Media Inform. Budidarma*, vol. 4, no. 3, pp. 855–862, 2020.
- [5] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017.
- [6] N. Azwanti, "Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 13, no. 1, p. 33, 2018.
- [7] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [8] S. Kasus, U. Dehasen, S. Haryati, A. Sudarsono, and E. Suryana, "IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4 . 5," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.