

MENGANALISA TINGKAT PENGANGGURAN DI DESA TITI MERAH DENGAN ALGORITMA C4.5

Roni Gustian Manik^{1*}, Agus Perdana Windarto², M.Fauzan³

^{1,2,3}STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: ¹ RonyGustian@gmail.com

Abstract

The purpose of this study was to analyze and find out the unemployment rate in Titi Merah Village, Kec. Lima Puluh, Kab. Batu Bara. By knowing the unemployment rate in the village, the Village Government can further evaluate to expand job vacancies for the unemployed. This study measures the unemployment rate in the village of red titi, Kec. Lima Puluh, Kab. Batu Bara. The method used in this study is C4.5 Algorithm, where the source of the data used is questionnaire / questionnaire technique that is given to the community randomly in the village of Titi Merah, Kec. Fifty, Kab.Coal. The variables used include (1) Competition in the World of Work, (2) Education Level, (3) Minimum Wages (4) Inflation and (5) Information. The results of this study are the most dominant factors causing the unemployment rate in the village of red titi and is expected to be input to the village administration as an evaluation material in tackling unemployment in the village of Titi Merah, Kec. Lima Puluh, Kab. Batu Bara.

Keywords: Unemployment, Data Mining Classification, Algorithm C4.5

Abstrak

Tujuan penelitian adalah untuk menganalisis dan mengetahui tingkat pengangguran di Desa Titi Merah, Kec. Lima Puluh, Kab. Batu Bara. Dengan mengetahui tingkat pengangguran yang ada di desa tersebut, pihak Pemerintahan Desa dapat mengevaluasi lebih lanjut untuk memperluas lowongan kerja bagi pengangguran. Penelitian ini melakukan pengukuran tingkat pengangguran di desa titi merah, Kec. Lima Puluh, Kab. Batu Bara. Metode yang digunakan dalam penelitian ini adalah Algoritma C4.5, dimana sumber data yang digunakan menggunakan teknik angket/kuisisioner yang diberikan kepada Masyarakat secara acak di lingkungan desa titi merah, Kec. Lima Puluh, Kab. Batu Bara. Adapun variabel yang digunakan diantaranya (1) Persaingan Dunia Kerja, (2) Tingkat Pendidikan, (3) Upah Minimum (4) Inflasi dan (5) Informasi. Hasil dari penelitian ini adalah di dapat nya faktor paling dominan penyebab tingkat pengangguran di desa titi merah dan di harapkan dapat menjadi masukan kepada pihak pemerintahan desa sebagai bahan evaluasi dalam menanggulangi pengangguran di Desa Titi Merah, Kec. Lima Puluh, Kab. Batu Bara.

Kata Kunci: Pengangguran, Data Mining, Klasifikasi, Algoritma C4.5

1. PENDAHULUAN

Pengangguran atau tuna karya adalah istilah untuk mereka yang tidak bekerja sama sekali, sedang mencari pekerjaan, bekerja kurang dari dua hari selama seminggu, atau seseorang yang sedang berusaha mendapatkan pekerjaan yang layak. Pengangguran umumnya disebabkan oleh meningkatnya jumlah angkatan kerja atau para pencari pekerjaan yang tidak sebanding dengan jumlah lapangan kerja yang tersedia di sebuah wilayah [1]. Dilansir dari pendapat jurnalis pikiran rakyat.com, Dhita Seftiawan 15 Agustus 2018 mengatakan bahwa setiap tahunnya sekolah menengah atas atau kerjuruan di Indonesia meluluskan ratusan bahkan ribuan siswa, namun hanya 20% dari jumlah kelulusan yang dapat melanjutkan pendidikan ke bangku perkuliahan, 30% dari mereka bekerja dan sisa nya menganggur. Angka pengangguran diIndonesia ini akan terus meningkat, dan menjadi masalah utama yang harus segera di tanggulangi oleh pemerintah.

Pengukuran tingkatpengangguran yang ada di sebuah wilayah adalah sebuah kegiatan yang seharusnya dilakukan secara berkala untuk mengetahui berapa tingkat pengangguran yang ada di wilayah tersebut. Jika tingkat pengangguran tinggi maka pemerintah harus mengevaluasi lebih lanjut

tentang permasalahan ini kemudian menyediakan lapangan pekerjaan untuk masyarakat yang menganggur. Tingkat pengangguran yang tinggi masuk kedalam masalah ekonomi dan sosial, para pengangguran suatu saat bisa kehilangan kepercayaan dirinya sehingga dapat mengakibatkan tindakan kriminal, perselisihan dengan masyarakat dan lain sebagainya.

Di Desa Titi Merah, Kec. Lima Puluh, Kab. Batu Bara, Provinsi Sumatera Utara belum diketahui nya faktor apa yang paling dominan pada tingkat pengangguran yang ada di desa tersebut. Oleh karena itu diperlukan nya pengukuran faktor tingkat pengangguran pada Desa Titi Merah. Dengan mengetahui faktor tingkat pengangguran yang ada, pihak pemerintahan desa dapat mengevaluasi lebih lanjut dan diharap memperluas lapangan pekerjaan untuk mengurangi tingkat pengangguran di desa tersebut. Menyikapi permasalahan tersebut, maka perlu untuk dilakukan analisis tingkat pengangguran di desa titi merah. Ada beberapa metode untuk mengetahui faktor tingkat pengangguran salah satu nya dengan pedekatan *Data Mining* menggunakan Algoritma jenis pohon keputusan (*decision tree*) yaitu Algoritma C4.5. Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif. Klasifikasi merupakan salah satu proses pada *data mining* yang bertujuan untuk menemukan pola yang berharga dari data yang berukuran relatif besar hingga sangat besar.

Data Mining yang juga dikenal sebagai *knowledge*, adalah salah satu bidang yang berkembang pesat karena besarnya kebutuhan akan nilai tambah dari tumpukan *database* skala besar yang terakumulasi sejalan dengan pertumbuhan teknologi informasi yang sangat pesat [2]. *Data Mining* disebut juga *Knowledge Discovery in Database* (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data [3]. Proses *Knowledge Discovery in Database* melibatkan hasil proses *data mining* (proses pengekstrakan kecenderungan suatu pola data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami. Kemudian hasil dari penelitian dilakukan pengujian dengan *software Rapid Miner* dengan menggunakan *apply model* dan *performance* [4].

Berdasarkan uraian di atas, Adapun variabel yang digunakan diantaranya yakni Persaingan Dunia Kerja, Tingkat Pendidikan, Upah Minimum, Inflasi dan Informasi. Penelitian ini diharapkan mampu mengetahui faktor dominan tingkat pengangguran yang ada di desa titi merah, kemudian hasilnya bisa dijadikan model bahan pengambilan keputusan di masa yang akan datang untuk pihak pemerintahan daerah setempat dalam menanggulangi pengangguran.

2. METODE PENELITIAN

2.1. Data Mining

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, *data mining* dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah relasional *database* yang besar. Kemampuan *Data mining* untuk mencari informasi bisnis yang berharga dari basis data yang sangat besar, dapat dianalogikan dengan penambangan logam mulia dari lahan sumbernya, teknologi dipakai untuk [5].

Dalam aplikasinya, *data mining* sebenarnya merupakan salah satu bagian proses *Knowledge Discovery in Database* (KDD) yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik. Adapun proses KDD sebagai berikut :

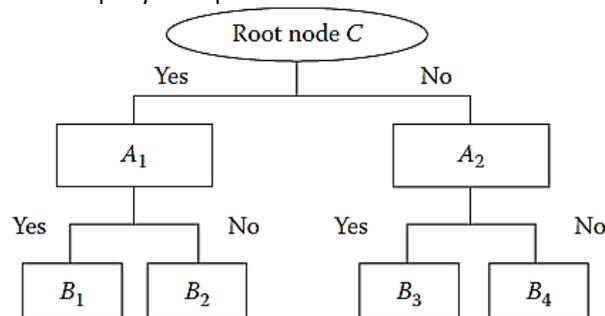
- a) *Data Selection*: pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalan informasi dalam KDD dimulai.
- b) *Preprocessing*: sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* dengan tujuan untuk membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.
- c) *Transformation*: yaitu proses *coding* pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam database.

- d) *Data mining*: proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.
- e) *Interpretation/ Evaluation*: pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya atau tidak.

2.2. Pohon Keputusan (Decision Tree)

Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) di mana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas [6]. Pada decision tree terdapat 3 jenis node, yaitu:

- a) Root Node, merupakan node paling atas, pada node ini tidak ada input dan bisa tidak mempunyai output atau mempunyai output lebih dari satu.
- b) Internal Node, merupakan node percabangan, pada node ini hanya terdapat satu input dan mempunyai output minimal dua.
- c) Leaf node atau terminal node, merupakan node akhir, pada node ini hanya terdapat satu input dan tidak mempunyai output.



Gambar 1. Contoh Struktur Decision Tree

2.3. Algoritma C4.5

Algoritma C.45 yaitu sebuah algoritma yang digunakan untuk membangun decision tree (pengambilan keputusan). Algoritma C.45 adalah salah satu algoritma induksi pohon keputusan, yaitu ID3 (Iterative Dichotomiser 3). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel training, label training, dan atribut. Algoritma C.45 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi missing value, bisa mengatasi kontinu data, dan pruning[7]. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C.45, yaitu :

- a) Menyiapkan data *training*. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas – kelas tertentu .
- b) Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing – masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus 1 (Craw, 2005).

$$Entropy(A) = \sum_{i=1}^n p_i * \log_2 p_i \tag{1}$$

Dimana :

S : Himpunan kasus

n : Jumlah partisi S

p_i : Proporsi dari S_i terhadap S

Setelah menghitung *entropy* setiap kasus selanjutnya menghitung nilai *gain* untuk pemisah objek dengan rumus 2 (Craw, 2005).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \tag{2}$$

Dimana :

S : Himpunan kasus

A : Atribut

N : Jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke-1

$|S|$: Jumlah kasus dalam S

- c) Ulangi langkah ke-2 hingga semua *record* terpartisi.
d) Proses partisi pohon keputusan akan berhenti saat :
- 1) Semua *record* dalam simpul N mendapat kelas yang sama.
 - 2) Tidak ada atribut di dalam *record* yang dipartisi lagi.
 - 3) Tidak ada *record* di dalam cabang yang kosong.

3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan di Desa Titi Merah, Kec. Lima Puluh, Kab. Batu Bara, Provinsi Sumatera Utara. Pada penelitian ini digunakan 5 kriteria dalam melakukan klasifikasi terhadap faktor tingkat pengangguran di Desa Titi Merah. Kriteria yang digunakan adalah sebagai berikut : (1) Persaingan Dunia Kerja, (2) Tingkat Pendidikan, (3) Upah Minimum (4) Inflasi dan (5) Informasi. Kuisisioner yang diperoleh selanjutnya dilakukan pencarian rata-rata dari setiap pertanyaan yang mewakili kriteria yang digunakan. Data yang digunakan merupakan jenis statistik *deskriptif* dengan masyarakat Desa Titi Merah.

Perhitungan Algoritma C4.5 untuk memperoleh model aturan pohon keputusan dapat diuraikan sebagai berikut :

Langkah 1: Menghitung jumlah kasus, jumlah kasus untuk Tingkat Pengangguran Tinggi, jumlah kasus untuk Tingkat Pengangguran Rendah dimana Persaingan Dunia Kerja = C1, Tingkat Pendidikan = C2, Upah Minimum = C3, Inflasi = C4 dan Informasi = C5.

Langkah 2: Menghitung *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan kelas atribut dengan persamaan (1). Selanjutnya dilakukan penghitungan *Gain* untuk masing-masing atribut dengan persamaan (2). Berikut ini adalah perhitungan nilai *entropy* dan *gain*.

Menghitung entropy total:

$$\text{Entropy [Total]} = \left(-\frac{78}{103} \times \log_2\left(\frac{78}{103}\right)\right) + \left(-\frac{25}{103} \times \log_2\left(\frac{25}{103}\right)\right) = 0,799531811$$

Langkah 3: Menghitung Entropy dan Gain untuk Node 1.

Menghitung entropy dan gain C1 :

$$\text{Entropy [C1-Sulit]} = \left(-\frac{28}{29} \times \log_2\left(\frac{28}{29}\right)\right) + \left(-\frac{1}{29} \times \log_2\left(\frac{1}{29}\right)\right) = 0,216396932$$

$$\text{Entropy [C1-Cukup Sulit]} = \left(-\frac{35}{40} \times \log_2\left(\frac{35}{40}\right)\right) + \left(-\frac{5}{40} \times \log_2\left(\frac{5}{40}\right)\right) = 0,543656443$$

$$\text{Entropy [C1-Tidak Sulit]} = \left(-\frac{15}{34} \times \log_2\left(\frac{15}{34}\right)\right) + \left(-\frac{19}{34} \times \log_2\left(\frac{19}{34}\right)\right) = 0,989992792$$

Information Gain [Total, C1] = 0,799531811

$$-\left(\left(\frac{29}{103} \times 0,216396932\right) + \left(\frac{40}{103} \times 0,543656443\right) + \left(\frac{34}{103} \times 0,989992792\right)\right) = 0,200717794$$

Menghitung entropy dan gain C2 :

$$\text{Entropy [C2-SD]} = \left(-\frac{39}{39} \times \log_2\left(\frac{39}{39}\right)\right) + \left(-\frac{0}{39} \times \log_2\left(\frac{0}{39}\right)\right) = 0$$

$$\text{Entropy [C2-SMP]} = \left(-\frac{39}{49} \times \log_2\left(\frac{39}{49}\right)\right) + \left(-\frac{10}{49} \times \log_2\left(\frac{10}{49}\right)\right) = 0,73001663$$

$$\text{Entropy [C2-SLTA]} = \left(-\frac{0}{15} \times \log_2\left(\frac{0}{15}\right)\right) + \left(-\frac{15}{15} \times \log_2\left(\frac{15}{15}\right)\right) = 0$$

Information Gain [Total, C2] = 0,799531811 - $\left(\left(\frac{39}{103} \times 0\right) + \left(\frac{49}{103} \times 0,73001663\right) + \left(\frac{15}{103} \times 0\right)\right) = 0,452242346$

Menghitung entropy dan gain C3 :

$$\text{Entropy [C3-Tinggi]} = \left(-\frac{7}{30} \times \log_2\left(\frac{7}{30}\right)\right) + \left(-\frac{23}{30} \times \log_2\left(\frac{23}{30}\right)\right) = 0,783776947$$

$$\text{Entropy [C3-Rendah]} = \left(-\frac{71}{73} \times \log_2\left(\frac{71}{73}\right)\right) + \left(-\frac{2}{73} \times \log_2\left(\frac{2}{73}\right)\right) = 0,181166402$$

$$\text{Information Gain [Total, C3]} = 0,799531811 - \left(\left(\frac{30}{103} \times 0,783776947\right) + \left(\frac{73}{103} \times 0,181166402\right)\right) = 0,44287774$$

Menghitung entropy dan gain C4 :

$$\text{Entropy [C4-Iya]} = \left(-\frac{43}{46} \times \log_2\left(\frac{43}{46}\right)\right) + \left(-\frac{3}{46} \times \log_2\left(\frac{3}{46}\right)\right) = 0,347816914$$

$$\text{Entropy [C4-Tidak]} = \left(-\frac{35}{57} \times \log_2\left(\frac{35}{57}\right)\right) + \left(-\frac{22}{57} \times \log_2\left(\frac{22}{57}\right)\right) = 0,962146133$$

$$\text{Information Gain [Total, C4]} = 0,799531811 - \left(\left(\frac{46}{103} \times 0,347816914\right) + \left(\frac{57}{103} \times 0,962146133\right)\right) = 0,1117463$$

Menghitung entropy dan gain C5 :

$$\text{Entropy [C5-Mudah]} = \left(-\frac{6}{20} \times \log_2\left(\frac{6}{20}\right)\right) + \left(-\frac{14}{20} \times \log_2\left(\frac{14}{20}\right)\right) = 0,88120899$$

$$\text{Entropy [C5-Cukup Mudah]} = \left(-\frac{11}{16} \times \log_2\left(\frac{11}{16}\right)\right) + \left(-\frac{5}{16} \times \log_2\left(\frac{5}{16}\right)\right) = 0,896038233$$

$$\text{Entropy [C5-Sulit]} = \left(-\frac{61}{67} \times \log_2\left(\frac{61}{67}\right)\right) + \left(-\frac{6}{67} \times \log_2\left(\frac{6}{67}\right)\right) = 0,43497348$$

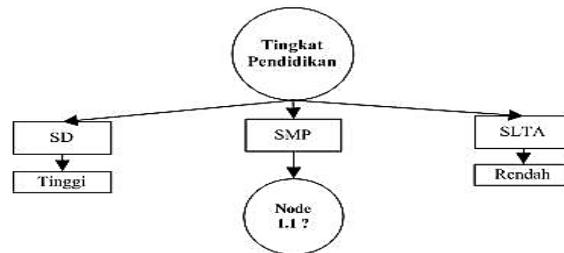
$$\text{Information Gain [Total, C5]} = 0,799531811 - \left(\left(\frac{20}{103} \times 0,88120899\right) + \left(\frac{16}{103} \times 0,896038233\right) + \left(\frac{67}{103} \times 0,43497348\right)\right) = 0,206273045$$

Berikut ini hasil perhitungan nilai entropy dan gain Node1 yang diuraikan pada Tabel.

Tabel 1. Hasil Perhitungan Node 1

Node 1	Jumlah Kasus (S)	Tinggi (S1)	Rendah (S2)	Entropy	Information Gain
TOTAL	103	78	25	0,799531811	
C1					0,200717794
Sulit	29	28	1	0,216396932	
Cukup Sulit	40	35	5	0,543564443	
Tidak Sulit	34	15	19	0,989992792	
C2					0,452242346
SD	39	39	0	0	
SMP	49	39	10	0,73001663	
SLTA	15	0	15	0	
C3					0,442847774
Tinggi	30	7	23	0,783776947	
Rendah	73	71	2	0,181166402	
C4					0,1117463
Iya	46	43	3	0,347816914	
Tidak	57	35	22	0,962146133	
C5					0,206273045
Mudah	20	6	14	0,881290899	
Cukup Mudah	16	11	5	0,896038233	
Sulit	67	61	6	0,43497348	

Dari hasil perhitungan pada Tabel 1. diperoleh nilai atribut tertinggi adalah Tingkat Pendidikan dengan Information gain sebesar 0,452242346. Maka atribut Tingkat Pendidikan dipilih sebagai node akar. Dari hasil perhitungan diatas dapat digambarkan pohon keputusan Node 1 yang ditunjukkan pada Gambar 4.1 berikut ini :



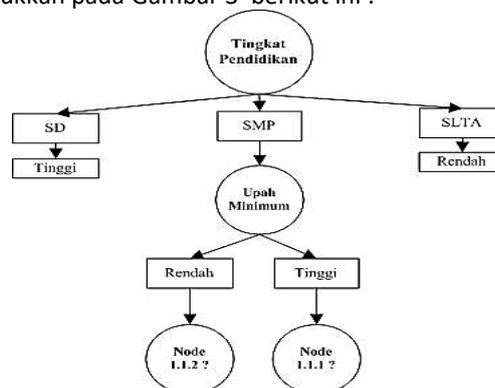
Gambar 2. Pohon Keputusan Hasil Perhitungan Node 1

Nilai kelas atribut SD mengklasifikasikan kasus menjadi satu keputusan yaitu Tingkat Pengangguran Tinggi. Nilai kelas atribut SLTA mengklasifikasikan kasus menjadi satu keputusan yaitu Tingkat Pengangguran Rendah. Untuk kelas atribut SMP belum diperoleh hasil antara keputusan Tingkat Pengangguran Tinggi ataupun Rendah, maka perlu dilakukan perhitungan lebih lanjut. Berikut ini hasil perhitungan dari kelas atribut C2–SMP ditunjukkan pada Tabel 2. berikut ini:

Tabel 2. Hasil Perhitungan Node 1.1

Node 1.1	Jumlah Kasus	Tinggi (S1)	Rendah (S2)	Entropy	Information Gain
C2-SMP	49	39	10	0,73001663	
C1					0,32718716
Sulit	17	10	7	0,977417818	
Cukup Sulit	17	10	5	0,969589482	
Tidak Sulit	15	5	4	1,036824992	
C3					0,623135832
Tinggi	13	4	9	0,89049164	
Rendah	36	35	1	0,183122068	
C4					0,496948373
Iya	21	20	1	0,276195428	
Tidak	28	19	9	0,905928216	
C5					0,512552595
Mudah	8	3	5	0,954434003	
Cukup Mudah	9	8	1	0,503258335	
Sulit	32	28	4	0,543564443	

Dari hasil perhitungan pada Tabel 2. diperoleh nilai atribut tertinggi dari C2 - SMP adalah C3 dengan Information gain sebesar 0,623135832. Dari hasil perhitungan diatas dapat digambarkan pohon keputusan Node 1.1 yang ditunjukkan pada Gambar 3 berikut ini :



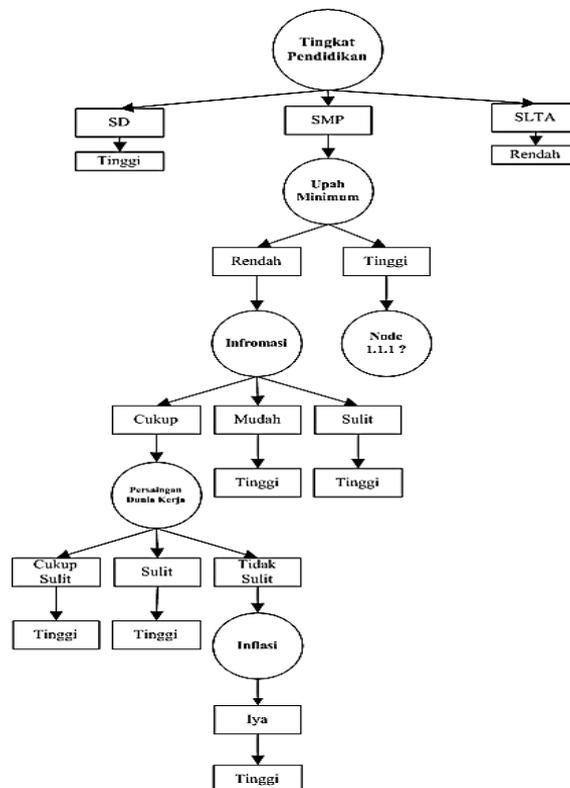
Gambar 3. Pohon Keputusan Hasil Perhitungan Node 1.1

Nilai kelas atribut Rendah dan Tinggi belum diperoleh hasil antara keputusan Tingkat Pengangguran Tinggi ataupun Rendah, maka perlu dilakukan perhitungan lebih lanjut. Proses perhitungan dilakukan hingga diperoleh hasil dengan node 1.1.2.1.1 dengan nilai kelas atribut Cukup Sulit dan Sulit mengklasifikasikan kasus menjadi satu keputusan yaitu Tingkat Pengangguran Tinggi. Untuk kelas atribut Tidak Sulit belum diperoleh hasil antara keputusan Tingkat Pengangguran Tinggi ataupun Rendah, maka perlu dilakukan perhitungan lebih lanjut. Kemudian hasil perhitungan untuk node 1.1.2.1.1C2 – SMP, C3- Rendah, C5 – Cukup Mudah, C1-Tidak Sulit ditunjukkan pada Tabel 3. berikut ini:

Tabel 3. Hasil Perhitungan Node 1.1.2.1.1

Node 1.1.2.1.1		Jumlah Kasus (S)	Tinggi (S1)	Rendah (S2)	Entropy	Information Gain
C2-SMP	C3 -Rendah	2	1	1	1	0,780114335
C5-Cukup Mudah						
C1-Tidak Sulit						
C4						
	Iya	2	1	1	1	
	Tidak	0	0	0	0	

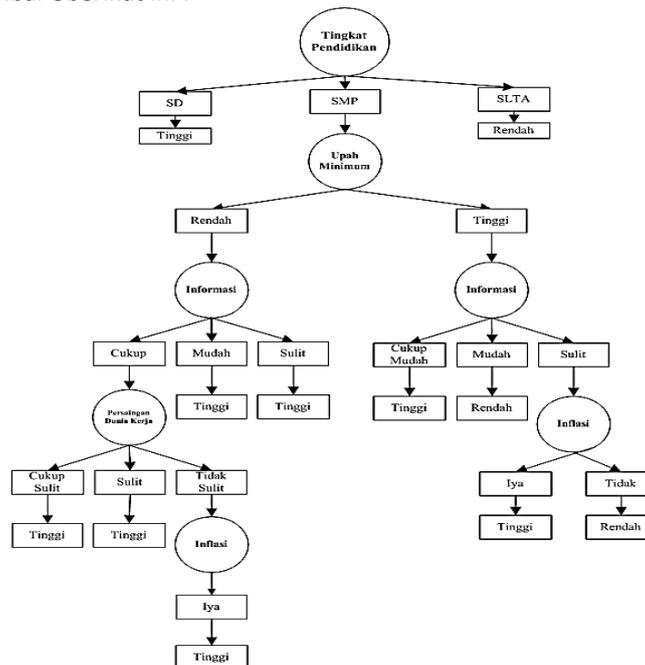
Dari hasil perhitungan pada Tabel 3. diperoleh nilai atribut tertinggi dari C2 – SMP, C3- Rendah, C5 – Cukup Mudah, C1-Tidak Sulit adalah C4 dengan Information gain sebesar 0,780114335. Dari hasil perhitungan diatas dapat digambarkan pohon keputusan Node 1.1.2.1.1 yang ditunjukkan pada Gambar 4 berikut ini :



Gambar 4. Pohon Keputusan Hasil Perhitungan Node 1.1.2.1.1

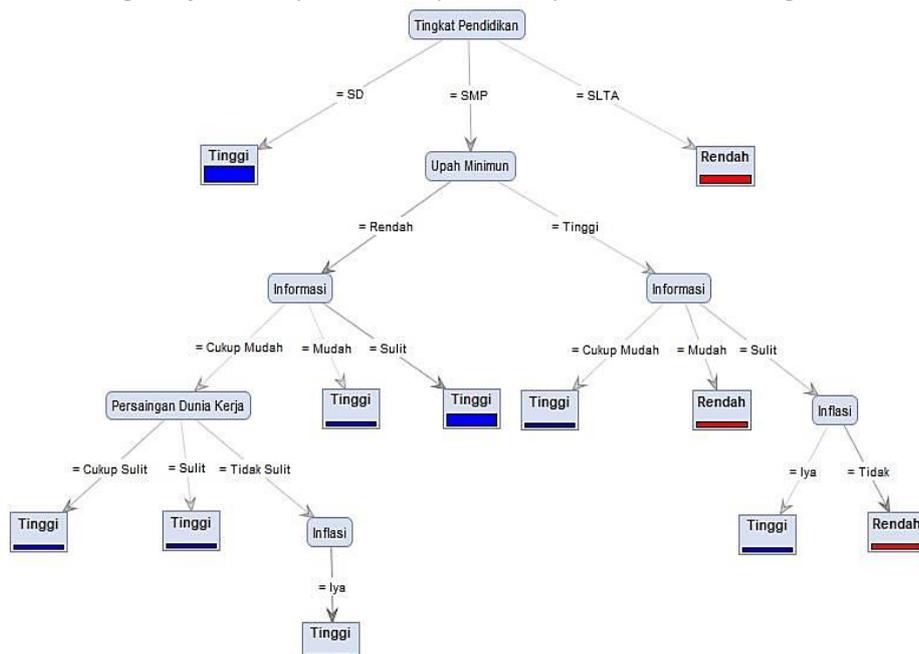
Nilai kelas atribut Iya mengklasifikasikan kasus menjadi satu keputusan yaitu Tingkat Pengangguran Tinggi, maka semua perhitungan telah selesai pada node ini sehingga tidak perlu dilakukan perhitungan lagi. Dengan demikian pohon keputusan pada gambar 4 merupakan pohon keputusan terakhir yang terbentuk. Kemudian pohon keputusan yang terbentuk dari keseluruhan

perhitungan dengan menggunakan algoritma C4.5 pada kasus tingkat pengangguran di desa titi merah ditunjukkan pada Gambar 5berikut ini :



Gambar 5. Pohon Keputusan Akhir Dari Hasil Perhitungan C4.5

Pada tahap akhir penerapan Algoritma C4.5 dilakukan penyesuaian hasil perhitungan manual melalui pengujian menggunakan *software RapidMiner 5.3*. Hasil pengolahan data dengan model pohon keputusan sesuai dengan *software RapidMiner*, dapat dilihat pada Gambar 6. sebagai berikut :



Gambar 6. Decision Tree Pada Rapidminer

Gambar 6. diatas merupakan pohon keputusan yang dihasilkan pada *Rapidminer* dengan aturan atau *rule* yang dapat dilihat pada text view pada Gambar 7. berikut :

-
- Data Mahasiswa Drop Out,” vol. 10, no. 2, pp. 2013–2016, 2017, doi: 10.3969/j.issn.1002-5006.2017.06.012.
- [3] Yulia and A. D. Putri, “Data Mining Menggunakan Algoritma C4 . 5 Untuk Memprediksi Kepuasan Mahasiswa Terhadap Terhadap Kinerja Dosen Di Kota Batam,” *Comput. Based Informaon Syst. J.*, vol. 07, pp. 56–66, 2019.
- [4] M. Yuli, “Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [5] A. G. Mabur, “Penerapan Data Mining Program Studi Teknik Informatika Jurnal Komputer dan Informatika (KOMPUTA),” 2012.
- [6] M. A. Sembiring, “Penerapan Metode Decision Tree Algoritma C45 Untuk Memprediksi Hasil Belajar Mahasiswa Berdasarkan Riwayat Akademik,” *J. Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 60–65, 2016.
- [7] 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, doi: 10.30872/jim.v13i1.629.