

Algoritma C4.5 Dalam Menentukan Arus Kas Pembangunan Jalan (Studi Kasus Di PT.Tasya Total Persada)

Nur Isnani

Universitas Putra Indonesia "YPTK" Padang

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 1 Februari 2018

Revisi Akhir: 10 Maret 2018

Diterbitkan Online: 23 Maret 2018

KATA KUNCI

Data Mining, Klasifikasi, Algoritma C4.5,
Pohon Keputusan, Proyek Konstruksi

KORESPONDENSI

No HP:

E-mail: nurisnani86@gmail.com

A B S T R A C T

PT Tasya Total Persada adalah sebuah Industri jasa konstruksi yang merupakan salah satu sektor industri yang memiliki Arus Kas pembangunan jalan. Keberhasilan dari suatu proyek dapat dilihat dari arus kasnya yang jelas dan terperinci untuk membantu manajer mengambil keputusan dalam proyek konstruksi. Arus kas dapat digunakan sebagai acuan yang dapat menilai kemampuan perusahaan secara rinci dan jelas untuk menilai kebutuhan perusahaan baik itu secara fisik atau non fisik. Arus kas dapat juga mempengaruhi kegiatan serta kinerja proyek yang sedang berlangsung, menjadikannya berhasil atau tidak. Metode yang digunakan dalam analisis ini adalah Algoritma C4.5 yang merupakan salah satu algoritma moderen untuk melakukan Data Mining. Data Mining adalah suatu proses pertambangan data yang digali meskipun belum diketahui hasilnya dan berikut diolah sedemikian rupa untuk mendapat hasil yang luar biasa. Algoritma C4.5 disebut juga dengan pohon keputusan (decision tree) yaitu merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon. Berdasarkan data yang dikumpulkan maka akan dicleaning untuk membuang data yang tidak konsisten dan noise sehingga dari data yang ada didapat beberapa atribut yang diambil yaitu Mobilisasi, Pekerjaan drainase, Pekerjaan tanah, Pekerjaan struktur, Biaya tenaga kerja, Material, Bina lingkungan. Setelah proses transformasi selesai maka dilanjutkan ke proses Data Mining dengan teknik klasifikasi dan menggunakan algoritma C4.5. Dengan terbentuknya pohon keputusan, dapat dilakukan analisa dan evaluasi rule yang dihasilkan. Rule inilah yang akan menjadi aturan dalam mengidentifikasi arus kas, sehingga manajer bisa mengambil keputusan dengan tepat dan cepat pada proyek konstruksi tersebut.

1. PENDAHULUAN

Dalam perkembangan industri jasa konstruksi sekarang ini, Keberhasilan dari suatu proyek dapat dilihat dari arus kasnya yang jelas dan terperinci. [1] Keberhasilan dari suatu proyek dapat dilihat dari arus kasnya yang jelas dan terperinci untuk membantu manajer mengambil keputusan dalam proyek industri konstruksi. Dengan adanya hal tersebut maka dalam penelitian ini akan dibuatkan sebuah Analisis Data Mining dengan menggunakan Algoritma C4.5.

Tujuan dari penelitian ini adalah sebagai berikut :

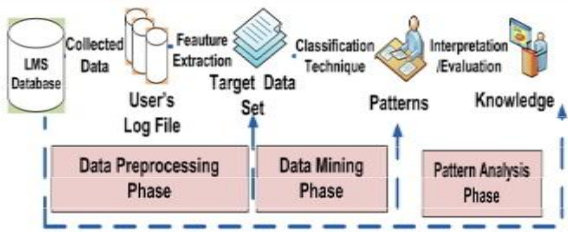
1. Mengetahui pengolahan data arus kas, sehingga manajer dapat mengambil keputusan dengan tepat dan cepat dibidang konstruksi pada PT.Tasya Total Persada.
2. Mengetahui bagaimana proses Algoritma C4.5.

3. Menganalisa data arus kas, sehingga manajer dapat mengambil keputusan dengan tepat dan cepat dengan menggunakan Algoritma C4.5
4. Mengimplementasikan Algoritma C4.5, sehingga manajer dapat mengambil keputusan dengan tepat dan cepa
5. Mengevaluasi setiap rule dengan membuat pohon keputusan.

2. TINJAUAN PUSTAKA

2.1 Knowledge Discovery In Database (KDD)

[2] Knowledge Discovery In Databases (KDD) Merupakan kegiatan yang meliputi tiga tahap utama yaitu: (a) data pre-processing; (B) data mining, dan (c) analisis pola. Dapat kita lihat dalam gambar berikut.

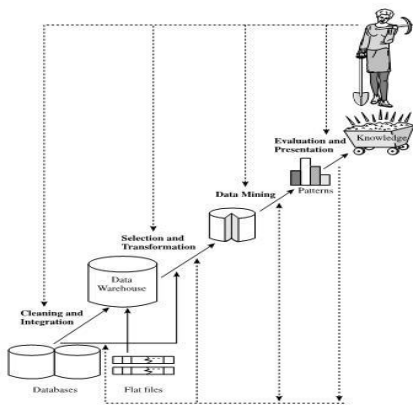


Gambar 1. Proses Knowledge Discovery In Databases (KDD)

- A. Tahap Data Pre-Processing
Proses mengumpulkan data, Data diambil dari beberapa bahagian kelompok yang akan dianalisa lalu diekstrak, Model dihasilkan dengan menggunakan kumpulan data berdasarkan pada turunan variabel.
- B. Tahap Data Mining
Suatu metode yang diterapkan untuk menemukan pengetahuan berharga yang tersembunyi dari data.
- C. Tahap Analisis Pola
Mengidentifikasi pola-pola menarik untuk dipresentasikan, selama yang dihasilkan belum sesuai maka langkah-langkahnya diulangi kembali.

2.2 Data Mining

Data Mining merupakan penambangan data yang terstruktur dimana kegiatan didalamnya meliputi seperti asosiasi, Clustering, Dan klasifikasi. [3]. Data Mining menurut [4] adalah suatu rangkaian proses yang dapat dibagi menjadi beberapa tahap.



Gambar 2. Tahap-Tahap Data Mining

Tahap-tahap tersebut yaitu :

1. Pemilihan data (untuk membuang data yang tidak konsisten dan noise)
2. Kombine data (penggabungan data dari beberapa sumber)
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk me-mining)
4. Aplikasi teknik Data Mining
5. Evaluasi pola yang ditemukan (untuk menemukan data yang menarik serta bernilai)
6. Presentasi pengetahuan (teknik visualisasi).

2.3 Algoritma C4.5

Secara umum algoritma C4.5 [5] untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar.

2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Algoritma C4.5 dapat menangani data numerik dan diskret. Algoritma C4.5 menggunakan rasio perolehan (gain ratio).

a. Konsep Entropi

Entropi (S) merupakan jumlah bit yang dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropi dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Entropi di gunakan untuk mengukur ketidak aslian S. Besarnya entropi pada ruang sampel S di definisikan dengan :

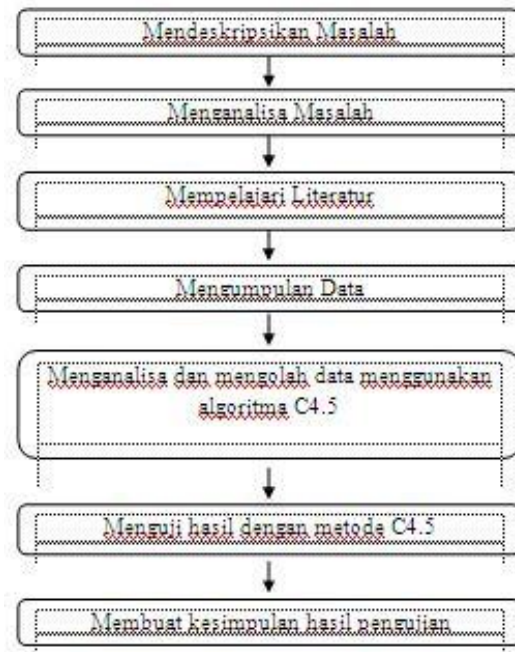
$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

b. Konsep Gain

Gain (S,A) merupakan Perolehan informasi dari atribut A relative terhadap output data S. Perolehan informasi didapat dari output data dependent S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S,A).

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

3. METODOLOGI



Gambar 3. Kerangka Kerja

Berdasarkan data yang dikumpulkan maka dilakukan proses seleksi, cleaning dan transformasi data sehingga ada beberapa atribut yang diambil yaitu mobilisasi, pekerjaan drainase, pekerjaan tanah, pekerjaan struktur, biaya tenaga kerja, material,bina lingkungan., Setelah proses transformasi selesai maka dilanjutkan ke proses data mining dengan teknik klasifikasi dan menggunakan algoritma C4.5, sehingga menghasilkan pohon keputusan dan aturan (rule).

4. HASIL DAN PEMBAHASAN

Penelitian ini akan pengolahan data arus kas, sehingga manajer dapat mengambil keputusan dengan tepat dan cepat dibidang konstruksi. yang menjadi *input* dalam penelitian ini adalah data inspeksi arus kas proyek konstruksi dan data kelengkapan lainnya sedangkan *outputnya* berupa faktor - faktor yang menyebabkan arus kas proyek konstruksi.
Output atau target variabel memiliki dua kategori yaitu baik dan tidak baik, maka digunakan teknik klasifikasi yang termasuk dalam pengelompokan *data mining*, karena dalam klasifikasi terdapat target variabel kategori.

4.1 Pengolahan Awal Data

Berdasarkan data yang dikumpulkan maka akan *dicleaning* untuk membuang data yang tidak konsisten dan *noise* sehingga dari data yang ada didapat beberapa atribut yang diambil yaitu mobilisasi, pekerjaan drainase, pekerjaan tanah, pekerjaan struktur, biaya tenaga kerja, material, bina lingkungan. Berdasarkan variabel-variabel yang sudah terpilih, *format* data berjumlah 12, maka akan mendapatkan 7 kelas, data tersebut akan dikelompokkan berdasarkan nilai atribut sebagai berikut :

Tabel 1 Klasifikasi Arus Kas

NILAI ATRIBUT	JUMLAH ATRIBUT	KLASIFIKASI
BAIK	≥ 4	BAIK
TIDAK BAIK	≥ 2	TIDAK BAIK
CUKUP	≥ 2	BAIK

Format data akhir setelah dilakukan pra-proses tampak seperti tabel 2 berikut ini:

Tabel 2 Data Setelah Pra-Proses

ESTIMASI ANGGARAN								
MOBILISASI	PEKERJAAN DRAINASE	PEKERJAAN TANAH	PEKERJAAN STRUKTUR	BIAYA TENAGA KERJA	MATERIAL	BINA LINGKUNGAN	TINGKAT CASH FLOW	
B	B	B	B	B	B	B	B	Baik
B	C	TB	B	TB	B	B	B	Tidak Baik
TB	TB	B	B	TB	TB	TB	TB	Tidak Baik
TB	C	B	TB	C	TB	B	B	Tidak Baik
C	TB	B	B	C	B	B	B	Baik
B	TB	TB	B	TB	B	B	B	Tidak Baik
C	B	TB	B	TB	B	TB	B	Tidak Baik
B	B	B	TB	C	B	B	B	Baik
B	TB	B	B	TB	TB	TB	TB	Tidak Baik
C	B	B	B	C	B	B	B	Baik
B	TB	B	TB	C	TB	B	B	Tidak Baik
C	B	B	B	C	B	TB	B	Baik

Format data akhir pada tabel 2 didapat berdasarkan dari *attribute* yang sudah dikelompokkan atau diklasifikasi, misalkan pada data Mobilisasi diklasifikasikan menjadi 3 yaitu “ Baik, Cukup dan Tidak Baik, pekerjaan drainase diklasifikasikan menjadi 2 yaitu “B dan TB “, begitupun dengan *attribute* lainnya.

4.2 Model Algoritma C4.5

Pada penelitian ini bertujuan untuk menganalisa data dengan menggunakan algoritma C4.5.

1. Pengujian Model

Pada pengujian tahap ini dilakukan pengujian model yaitu menghitung dan mendapatkan rule-rule yang ada pada model algoritma C4.5 yang diusulkan. Menentukan atribut sebagai akar dan menghitung nilai informasi gain atribut. Untuk memilih

atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada, Dibutuhkan nilai Entropy untuk menentukan gain tertinggi. Menghitung Nilai Entropy tiap-tiap atribut:

Entropy (Total) dengan rumus sebagai berikut :

$$Entropy (total) = - \sum_{i=1}^n - P_i \times \log_2 P_i$$

$$= \left(-\frac{5}{12} * \log_2 \left(\frac{5}{12} \right) \right) + \left(-\frac{7}{12} * \log_2 \left(\frac{7}{12} \right) \right) = 0,9799$$

Entropy (total) adalah menghitung nilai total keputusan Baik (5) dan tidak baik (7), sedangkan 12 adalah jumlah keseluruhan kasus. Kemudian hitung nilai entropy pada masing-masing atribut, sebagai contoh dapat kita lihat pada atribut Mobilisasi.

$$Entropy(B) = \left(-\frac{2}{6} * \log_2 \left(\frac{2}{6} \right) \right) + \left(-\frac{4}{6} * \log_2 \left(\frac{4}{6} \right) \right) = 0,9182$$

$$Entropy(C) = \left(-\frac{3}{4} * \log_2 \left(\frac{3}{4} \right) \right) + \left(-\frac{1}{4} * \log_2 \left(\frac{1}{4} \right) \right) = 0,8113$$

$$Entropy(TB) = \left(-\frac{0}{2} * \log_2 \left(\frac{0}{2} \right) \right) + \left(-\frac{2}{2} * \log_2 \left(\frac{2}{2} \right) \right) = 0$$

Menghitung Nilai Gain tiap-tiap atribut : Contohnya Gain (Total, Mobilisasi)

$$= 0,9799 - \left(\left(\frac{6}{12} * 0,9183 \right) + \left(\frac{4}{12} * 0,8113 \right) + \left(\frac{2}{12} * 0 \right) \right) = 0,2503$$

Dengan pembahasan di atas dapat dilihat hasilnya pada table 3

Table 3 Perhitungan Node 1

Node 1	Jumlah Anggaran (S)	BAIK(S1)	Tidak BAIK(S2)	Entropy	Gain
Total	12	5	7	0,979868757	
MOBILISASI					
BAIK	B	6	2	4	0,918295834 0,250294798
CUKUP	C	4	3	1	0,811278124
TIDAK BAIK	TB	2	0	2	0
PEKERJAAN DRAINASE					
BAIK	B	5	4	1	0,721928095 0,378262011
CUKUP	C	2	0	2	0
TIDAK BAIK	TB	5	1	4	0,721928095
PEKERJAAN TANAH					
BAIK	B	9	5	4	0,991076637 0,236561279
TIDAK BAIK	TB	3	0	3	0
PEKERJAAN STRUKTUR					
BAIK	B	9	4	5	0,991076637 0,466135328
TIDAK BAIK	TB	3	1	2	0,918296195
BIAYA TENAGA KERJA					
BAIK	B	1	1	0	0 1,439016674
CUKUP	C	6	4	2	0,918295834
TIDAK BAIK	TB	5	0	5	0
MATERIAL					
BAIK	B	8	5	3	0,954434003 0,343579421
TIDAK BAIK	TB	4	0	4	0
BINA LINGKUNGAN					
BAIK	B	8	4	4	1 0,583628131
TIDAK BAIK	TB	4	1	3	0,811278124

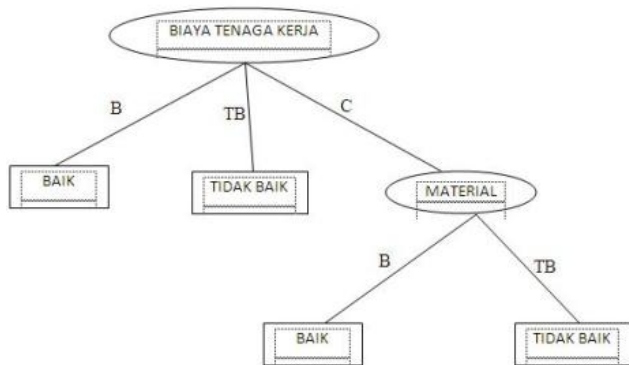
Dari perhitungan pada tabel 3 dapat diketahui bahwa atribut dengan gain tertinggi adalah Biaya tenaga kerja dengan Nilai 1.4390, atribut tersebut dapat menjadi node akar, di mana nilai

atribut B sudah dapat di katakana Baik dan TB sudah dapat dikatakan Tidak Baik. Akan tetapi, nilai atribut c masih perlu dilakukan perhitungan lagi. Pada table 4 dapat dilihat perhitungan node 1.1.

Table 4 Perhitungan Node 1.1

NODE 1.1		Jumlah Anggaran (S)	BAIK(S1)	TIDAK BAIK (S2)	Entropy	Gain
Node 1.1	Total Biaya Tenaga Kerja (Cukup)	6	4	2	0.918295834	
MOBILISASI						
	BAIK	B	2	1	1	0.584982501
	CUKUP	C	3	3	0	0
	TIDAK BAIK	TB	1	0	1	0
PEKERJAAN DRAINASE						
	BAIK	B	3	3	0	0.584982501
	CUKUP	C	1	0	1	0
	TIDAK BAIK	TB	2	1	1	1
PEKERJAAN TANAH						
	BAIK	B	6	4	2	0.918295834
	TIDAK BAIK	TB	0	0	0	0
PEKERJAAN STRUKTUR						
	BAIK	B	3	3	0	0.459147917
	TIDAK BAIK	TB	3	1	2	0.918295834
MATERIAL						
	BAIK	B	4	4	0	0.918295834
	TIDAK BAIK	TB	2	0	2	0
BINA LINGKUNGAN						
	BAIK	B	5	3	2	0.970950594
	TIDAK BAIK	TB	1	1	0	0.109170339

Dari hasil tabel 4 dapat diketahui bahwa atribut gain tertinggi adalah Material, dengan nilai 0.9183, berarti Material dapat menjadi node akar selanjutnya, Atribut B memiliki Nilai “Baik” Atribut TB bernilai “ Tidak Baik “, Pohon keputusan yang terbentuk pada node 1.1 dapat dilihat pada gambar 3



Gambar 5 Pohon Keputusan

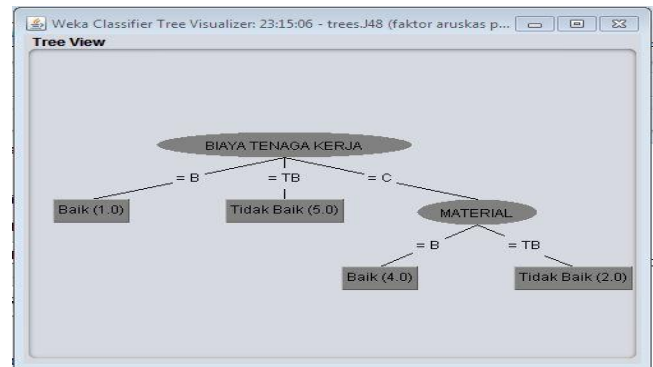
Adapun aturan atau rule yang terbentuk berdasarkan pohon keputusan terakhir seperti pada gambar 5 di atas adalah sebagai berikut :

1. IF Biaya Tenaga Kerja = Tidak Baik THAN Arus Kas = Tidak Baik.
2. IF Biaya Tenaga Kerja = Baik THAN Arus Kas = Baik
3. IF Biaya Tenaga Kerja = Cukup AND Material = Baik THAN Arus Kas = Baik
4. IF Biaya Tenaga Kerja = Cukup AND Material = Tidak Baik THAN Arus Kas = Tidak Baik.

Berdasarkan dari rule/knowledge yang dihasilkan terdapat beberapa rules cukup sesuai dengan kejadian yang terjadi terdiri dari 2 kasus utama yang sering terjadi yaitu :

- a. Biaya tenaga kerja
- b. Material

Dengan pengujian weka 3.8 dapat kita lihat kesamaan hasil pada gambar 4



Gambar 4 Hasil aplikasi weka 3.8

Dari tree view oleh WEKA dapat dilihat rule atau aturan yang terbentuk adalah sebagai berikut :

1. IF Biaya Tenaga Kerja = Tidak Baik THAN Arus Kas = Tidak Baik.
2. IF Biaya Tenaga Kerja = Baik THAN Arus Kas = Baik
3. IF Biaya Tenaga Kerja = Cukup AND Material = Baik THAN Arus Kas = Baik

IF Biaya Tenaga Kerja = Cukup AND Material = Tidak Baik THAN Arus Kas = Tidak Baik

5. KESIMPULAN DAN SARAN

Dengan uraian bab-bab yang sudah dibahas sebelumnya, kita dapat mengambil beberapa kesimpulan, yaitu :

1. Pemilihan variabel (atribut kondisi dan atribut keputusan) yang digunakan dalam mengklasifikasikan sangat mempengaruhi rule atau knowledge yang dihasilkan.
2. Algoritma C4.5 dapat membantu dalam mengolah serta mengimplementasikan data arus kas, sehingga manajer dapat mengambil keputusan dengan tepat dan cepat pada proyek konstruksi.
3. Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi rule prediksi untuk menggambarkan proses yang terkait dengan mengolah serta mengimplementasikan data arus kas, sehingga manajer dapat mengambil keputusan dengan tepat dan cepat.
4. Pengujian dengan WEKA lebih efektif dan fleksibel karena pohon keputusan yang terbentuk hampir sama dengan perhitungan yang dilakukan dengan manual. Oleh karena itu, penggunaan WEKA juga sangat membantu dalam pengujian yang dilakukan.

DAFTAR PUSTAKA

[1] G. Waligora and R. Rozycki, “Heuristic solving some discrete-continuous project scheduling problems with discounted cash flows,” 2016 21st Int. Conf. Methods Model. Autom. Robot. MMAR 2016, no. 2013, pp. 971–974, 2016.

[2] B. E. V Comendador, L. W. Rabago, and B. T. Tanguilig, “An educational model based on Knowledge Discovery in Databases (KDD) to predict learner’s behavior using classification techniques,” 2016 IEEE Int. Conf. Signal Process.

Commun. Comput., pp. 1–6, 2016.

- [3] M. Nunez-del-Prado and R. Gomez, “Learning data analytics through a Problem Based Learning course,” *2017 IEEE World Eng. Educ. Conf.*, pp. 52–56, 2017.
- [4] G. Indrawan and G. R. Dantes, “Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Stmik Denpasar Menggunakan Metode Technique For Others Reference By Similarity To Ideal Solution Program Studi Ilmu Komputer , Program Pascasarjana,” vol. 5, no. 2, pp. 746–760, 2016.
- [5] W. Supriyanti, Kusriani, and A. Amborowati, “Perbandingan kinerja algoritma c4.5 dan naive bayes untuk ketepatan pemilihan konsentrasi mahasiswa,” *Informa*, vol. 1, no. 2012, pp. 46–52, 2016.

BIODATA PENULIS



Nur Isnani, Lahir di Siberut, 26 Mei 1986. Meraih gelar sarjana komputer (S.Kom) dari Universitas Putra Indonesia "YPTK" Padang pada tahun 2008. Kemudian Gelar Magister Komputer (M.Kom) dari Univeritas Putera Indonesia “YPTK” Padang pada tahun 2016. Penulis dapat di hubungi melalui email nurisnanani86@gmail.com.