



# Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia

Tukino

Jurusan Sistem Informasi, Fakultas Teknik dan Komputer, Universitas Putera Batam

*Naskah Diterima : 10 November 2018; Diterima Publikasi : 29 April 2019*

*DOI : 10.21456/vol9iss1pp39-46*

---

## Abstract

In the construction project activities, planning is used as a reference for job implementers and becomes the standard of project implementation, including: documents, technical specifications, schedule and budget , for the purpose of this study was to determine how the process of Data Mining methods Algorithms C4.5 can help in seeing a pattern of budget cost planning consultant at PT SMOE Indonesia Batam in predicting Corporate profits as well as to analyze and evaluate the rule produced from decision tree (decision tree) in order to perform a prediction of the profits of the company at PT SMOE Indonesia, Batam. Inappropriate planning, inaccurate project realization investigations, inadequate project management skills and lack of professional service providers, are closely related to the outcome of a construction project process. PT SMOE Indonesia which is a company engaged in construction consulting services. At the present time PT SMOE Indonesia has done many construction planning projects both from government and private, this research will discuss how data mining with algorithm C4.5 process data from budget plan consultant planner cost to predict company profit. Data mining is a technique for extracting new information from piles or data warehouses, as we know information is seen as something that is very important and valuable because by mastering information it is easy to achieve a desired goal, this makes everyone race to while C4.5 algorithm is one of induction algorithm of decision tree that is ID3 (Iterative Dichotomiser 3). ID3 was developed by J. Ross Quinlan. In the ID3 algorithm procedure, the inputs are training samples, training labels and attributes. which will illustrate the profit prediction, the results of this study will result in the rules of profit and loss decisions company.

**Keywords** : Profit; Data Mining; Algorithm C4.5; Tree Decision.

## Abstrak

Dalam kegiatan proyek konstruksi, perencanaan dipergunakan sebagai bahan acuan bagi pelaksana pekerjaan dan menjadi standar pelaksanaan proyek, meliputi: dokumen, spesifikasi teknik, jadwal dan anggaran, untuk itu tujuan dari penelitian ini adalah untuk mengetahui bagaimana proses Data Mining dengan metode Algoritma C4.5 dapat membantu dalam melihat pola anggaran biaya konsultan perencana pada PT SMOE Indonesia Batam dalam memprediksi keuntungan Perusahaan serta untuk menganalisa dan mengevaluasi *rule* yang dihasilkan dari pohon keputusan (*decision tree*) guna melakukan prediksi keuntungan perusahaan pada PT SMOE Indonesia Batam. Perencanaan yang tidak tepat, investigasi realisasi proyek yang tidak sempurna, kurang memadainya kemampuan pengelolaan proyek dan kurang profesionalnya penyedia jasa, berkaitan erat terhadap hasil suatu proses proyek konstruksi. PT SMOE Indonesia yang merupakan sebuah perusahaan yang bergerak dalam bidang jasa konsultan konstruksi. Pada saat sekarang ini PT SMOE Indonesia telah banyak mengerjakan proyek perencanaan konstruksi baik dari pemerintah maupun swasta, penelitian ini akan membahas bagaimana data mining dengan algoritma C4.5 mengolah data-data dari rencana anggaran biaya konsultan perencana untuk memprediksi profit perusahaan. Data mining merupakan sebuah teknik untuk menggali informasi baru dari tumpukan atau gudang data, sebagaimana yang kita ketahui informasi dipandang sebagai sesuatu hal yang sangat penting dan berharga karena dengan menguasai informasi maka dengan mudah untuk mencapai sebuah tujuan yang diinginkan, hal ini membuat setiap orang berlomba untuk memperoleh informasi. Sedangkan algoritma C4.5 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. yang akan menggambarkan prediksi profit, hasil dari penelitian ini akan menghasilkan rule-rule keputusan profit dan kerugian perusahaan.

**Keywords** : Keuntungan; Data Mining; Algoritma C4.5; Pohon Keputusan.

---

\*) Penulis korespondensi: [tukino@gmail.com](mailto:tukino@gmail.com)

## 1. Pendahuluan

Tidak dapat dipungkiri lagi teknologi dan informasi ternyata membawa perubahan dalam segala bidang termasuk kepada perkembangan dunia bisnis. Perkembangan teknologi yang pesat saat ini banyak menarik perhatian masyarakat luas. Demikian pula pada perkembangan komputer yang dinamis diiringi perkembangan zaman dapat menghasilkan informasi yang cepat dan akurat dan tidak memerlukan waktu yang lama dalam menggali informasi yang dibutuhkan, termasuk pada perusahaan-perusahaan yang bergerak dalam jasa konsultan konstruksi dimana kegiatan usahanya yang selalu membutuhkan perhitungan yang signifikan terhadap informasi untuk menjalankan sebuah proyek perencanaan yang mereka kelola setiap tahunnya.

Dalam kegiatan proyek konstruksi, perencanaan dari jasa konsultan dipergunakan sebagai bahan acuan bagi pelaksana pekerjaan dan menjadi standar pelaksanaan proyek, meliputi: dokumen, spesifikasi teknik, jadwal dan anggaran. Perencanaan yang tidak tepat, investigasi realisasi proyek yang tidak sempurna, kurang memadainya kemampuan pengelolaan proyek dan kurang profesionalnya penyedia jasa, berkaitan erat terhadap hasil suatu proses proyek konstruksi, salah satunya dikota batam ini adalah PT SMOE Indonesia yang merupakan sebuah perusahaan yang bergerak dalam bidang jasa konsultan konstruksi. Pada saat sekarang ini PT SMOE Indonesia telah banyak mengerjakan proyek perencanaan konstruksi baik dari pemerintah maupun swasta, sebelum mereka mendapatkan proyek tersebut ada suatu langkah yang harus dilakukan yaitu mengikuti lelang yang diadakan oleh pemerintah atau swasta sebagai owner proyek, dalam hal ini tentunya PT SMOE Indonesia akan mempersiapkan berkas-berkas sebagai syarat mengikuti lelang tersebut termasuk dengan anggaran yang dibutuhkan selama melaksanakan proyek tersebut.

Berdasarkan Surat Edaran Menteri Pekerjaan Umum dan Perumahan Rakyat No.01/SE/M/2017, tentang Penentuan Biaya Langsung Personil (*Remuneration/ Billing Rate*) Dalam Penyusunan Harga Perkiraan Sendiri Pengadaan Jasa Konsultansi Konstruksi di Lingkungan Kementerian Pekerjaan Umum dan Perumahan Rakyat, pada Butir E menyebutkan bahwa penentuan besaran Biaya Langsung Personil (*Remuneration/Billing Rate*) dalam penyusunan HPS mempertimbangkan informasi biaya satuan yang dipublikasikan secara resmi oleh asosiasi terkait dan sumber data lain yang dapat dipertanggungjawabkan. Selama ini perusahaan memiliki sekumpulan data terkait pencapaian kinerja perusahaan yang telah menghabiskan banyak biaya namun data tersebut belum dapat dimanfaatkan dengan baik, contohnya saja pada saat menyusun anggaran biaya perusahaan kadang tidak memikirkan keuntungan atau profit yang akan mereka dapatkan,

karena mereka hanya berfokus untuk menang pada lelang pekerjaan saja, sehingga pada saat mengerjakan proyek yang telah mereka dapatkan sering terjadi kerugian atau tumpang tindih biaya operasional yang berujung pada kerugian.

Selama ini perusahaan memiliki sekumpulan data terkait pencapaian kinerja perusahaan yang telah menghabiskan banyak biaya namun data tersebut belum dapat dimanfaatkan dengan baik, contohnya saja pada saat menyusun anggaran biaya perusahaan kadang tidak memikirkan keuntungan atau profit yang akan mereka dapatkan, karena mereka hanya berfokus untuk menang pada lelang pekerjaan saja, sehingga pada saat mengerjakan proyek yang telah mereka dapatkan sering terjadi kerugian atau tumpang tindih biaya operasional yang berujung pada kerugian.

Berdasarkan masalah diatas melalui penerapan teknologi data mining dengan algoritma C4.5 maka sekumpulan data anggaran biaya yang belum dimanfaatkan tersebut akan diproses dan menghasilkan *new knowlage* yang menjadi sesuatu yang berharga dan berguna bagi perusahaan terutama dalam peningkatan profit. Data Mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut, untuk itu tujuan dari penelitian ini adalah untuk mengetahui bagaimana proses Data Mining dengan metode Algoritma C4.5 dapat membantu dalam melihat pola anggaran biaya konsultan perencanaan pada PT SMOE Indonesia Batam dalam memprediksi keuntungan Perusahaan serta untuk menganalisa dan mengevaluasi *rule* yang dihasilkan dari pohon keputusan (*decision tree*) guna melakukan prediksi keuntungan perusahaan pada PT SMOE Indonesia Batam.

## 2. Kerangka Teori

### 2.1. Profit

*Profit* atau disebut juga dengan laba menurut Martono dan Harijito menjelaskan bahwa konsep laba merupakan konsep yang menghubungkan Antara pendapatan atau penghasilan yang diperoleh oleh perusahaan di satu pihak, dan biaya yang harus ditanggung atau dikeluarkan oleh pihak lain. Untung atau laba didefinisikan sebagai kenaikan modal saham dari transaksi yang bersifat insidental dan bukan merupakan kegiatan pokok perusahaan dan dari transaksi lainnya yang mempengaruhi perusahaan dalam periode tertentu. Sedangkan profit bersih atau laba bersih merupakan laba dari bisnis perusahaan yang sedang berjalan setelah bunga dan pajak (Wowor dan Mangantar, 2012).

### 2.2. Data Mining

*Data mining*, sering juga disebut sebagai *knowledge discovery in database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian

data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso *et al.*, 2016). *knowledge discovery in database* (KDD) pada intinya adalah proses menemukan pengetahuan yang bermanfaat dari kumpulan data. Berstein *et al.* (2005) juga mendefinisikan *knowledge discovery in database* (KDD) sebagai hasil proses penjelajahan yang melibatkan penerapan berbagai Prosedur algoritma untuk memanipulasi data, membangun model dari data, dan memanipulasi model (Gamarra *et al.*, 2016).

Menurut Turban dalam bukunya yang berjudul "Decision Support Systems and Intelligent Systems", *data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar (Gunadi dan Sensuse, 2012).

### 2.2.1. Klasifikasi

Menurut Tan *et al* (2009), klasifikasi adalah proses untuk menyatakan suatu objek ke salah satu kategori yang sudah didefinisikan sebelumnya. Klasifikasi juga bisa didefinisikan sebagai proses pembelajaran fungsi target (model klasifikasi) yg memetakan setiap sekumpulan atribut  $x$  (input) ke salah satu kelas  $y$  yang didefinisikan sebelumnya (Zulkifli, 2016).

Lebih lanjut, *input* didefinisikan sebagai sekumpulan *record* (*training set*), dan setiap *record* terdiri atas sekumpulan atribut, salah satu atribut adalah kelas. Adapun model klasifikasi digunakan untuk antara lain :

1. Pemodelan Deskriptif sebagai perangkat penggambaran untuk membedakan objek-objek dari kelas berbeda.
2. Pemodelan Prediktif digunakan untuk memprediksi label kelas untuk *record* yang tidak diketahui atau tidak dikenal.

### 2.2.2. Pohon Keputusan

Manusia selalu dihadapkan dengan berbagai macam masalah dari berbagai bidang kehidupan. Masalah ini juga memiliki variasi tingkat kesulitannya. Untuk menghadapi masalah ini manusia mulai mengembangkan sebuah sistem untuk membantu mereka menyelesaikan masalah-masalah ini, salah satu sistem tersebut adalah pohon keputusan.

Pohon keputusan adalah metode klasifikasi dan prediksi yang sudah terbukti *powerfull* dan sangat terkenal. Metode ini berfungsi untuk mengubah fakta menjadi pohon keputusan yang merepresentasikan aturan yang dapat mudah dimengerti dengan bahasa alami. Proses dari pohon keputusan ini dimulai dari *node* akar hingga *node* daun yang dilakukan secara rekursif dimana setiap percabangan menyatakan

kondisi dan setiap ujung pohon akan menyatakan keputusan (Haryati *et al.*, 2015)

*Decision tree* merupakan metode yang mengubah data menjadi pohon keputusan (*decision tree*) dan aturan-aturan keputusan (Thi *et al.*, 2015), 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*.

### 2.2.3. Algoritma C4.5

Algoritma C4.5 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 C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi *continuu data*, dan *pruning* (Faradillah, 2013).

Sebuah objek yang diklasifikasikan dalam pohon harus dites nilai *Entropy*-nya. *Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* dari kumpulan data. Dari nilai *Entropy* tersebut kemudian dihitung nilai *information gain* (IG) masing-masing atribut. *Entropy* ( $S$ ) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel  $S$ . *Entropy* dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai *Entropy* maka akan semakin *Entropy* digunakan dalam mengekstrak suatu kelas. *Entropy* digunakan untuk mengukur ketidaksiharian  $S$ .

Untuk memilih atribut akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus menghitung *Gain* seperti yang tertera dalam persamaan berikut:

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

Di mana :

- $S$  : himpunan kasus
- $A$  : atribut
- $N$  : jumlah partisi atribut  $A$
- $|S_i|$  : jumlah kasus pada partisi ke- $i$
- $|S|$  : jumlah kasus dalam  $S$

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan 2 berikut.

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2)$$

Di mana :

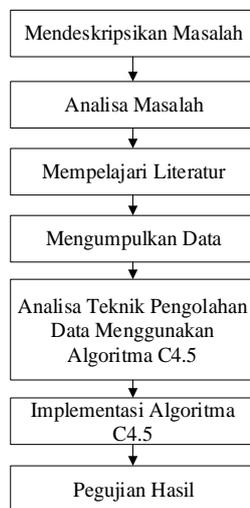
- $S$  : himpunan kasus  
 $A$  : fitur  
 $N$  : jumlah partisi  $S$   
 $p_i$  : proporsi dari  $S_i$  terhadap  $S$

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Lestari dan Suryadi, 2014):

- Pilih atribut sebagai akar.
- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

### 3. Metode

Pada bagian ini akan diuraikan desain penelitian, desain penelitian ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Adapun desain penelitian ini dapat dilihat pada Gambar 1:



Gambar 1. Desain penelitian

Berdasarkan desain penelitian pada gambar 1, maka masing-masing langkahnya dapat diuraikan seperti berikut ini :

#### 1. Mendeskripsikan Masalah

Mendeskripsikan masalah yang akan teliti perlu ditentukan terlebih dahulu. Mendeskripsikan masalah dalam penelitian dengan menentukan dan mendefinisikan batasan masalah yang akan diteliti, sehingga membantu dalam mendapatkan suatu solusi yang terbaik dari masalah tersebut. Jadi, langkah pertama ini adalah langkah awal yang terpenting dalam penelitian ini.

#### 2. Analisa Masalah

Langkah analisis masalah merupakan langkah untuk dapat memahami masalah yang telah ditentukan ruang lingkup atau batasannya. Dengan

menganalisis masalah yang telah ditentukan tersebut, maka diharapkan masalah dapat dipahami dengan baik.

#### 3. Mempelajari Literatur

Untuk mencapai tujuan, maka dipelajari beberapa literatur-literatur yang diperkirakan dapat digunakan. Kemudian literatur-literatur yang dipelajari tersebut diseleksi untuk dapat ditentukan literatur-literatur mana yang akan digunakan dalam penelitian.

#### 4. Mengumpulkan Data

Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara langsung di minimarket ayu sehingga permasalahan yang ada dapat diketahui secara jelas. Kemudian dilakukan *interview* yang bertujuan untuk mendapatkan informasi atau data yang dibutuhkan. Selain itu juga dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam melakukan analisis terhadap data dan informasi yang didapat. Analisa teknik pengolahan data menggunakan algoritma C4.5. Data yang diperoleh dari tempat penelitian selanjutnya dilakukan analisa dan pengolahan menggunakan algoritma C4.5.

#### 5. Perancangan Algoritma C4.5

Pada tahap ini akan dilakukan proses perancangan dari model sistem dengan algoritma C4.5 sehingga membentuk pohon keputusan (*decision tree*) dan menghasilkan suatu *rule* prediksi Profit Perusahaan.

#### 6. Implementasi Algoritma C4.5, dengan langkah-langkah sebagai berikut :

- Menentukan atribut sebagai akar dan mengitung nilai informasi *gain* atribut.
- Menyusun *Tree* awal
- Mengubah *Tree* menjadi *rule*

#### 7. Pengujian Hasil

Pada tahap ini, penulis melakukan pengujian dan hasil perancangan sistem menggunakan *software Data Mining open source WEKA*. Sistem diuji dengan prosedur-prosedur untuk melakukan eksplorasi dan permodelan dari data-data yang ada sehingga mendapatkan suatu hubungan tersembunyi dari data tersebut.

Penelitian ini memakai beberapa metode guna mendukung pengambilan data pada objek peneitian dengan mendatangi dan menghubungi langsung PT SMOE Indonesia, metode yang dilakukan adalah :

#### 1. Wawancara

Wawancara adalah teknik pengumpulan data yang dilakukan melalui tatap muka dan tanya jawab langsung antara peneliti dan narasumber.pada metode ini peneliti langsung menyanyakan beberapa hal yang tentang kebutuhan informasi berkaitan dengan perusahaan terutama pada bidang anggaran proyek.

#### 2. Observasi

Metode pengumpulan data yang kompleks karena melibatkan berbagai faktor dalam pelaksanaannya. Metode pengumpulan data observasi tidak hanya mengukur sikap dari responden, namun juga dapat digunakan untuk merekam berbagai fenomena yang terjadi dengan melihat langsung agar memperoleh data yang objektif.

3. Studi Dokumen/Perpustakaan

Kegiatan untuk menghimpun informasi yang relevan dengan topik atau masalah yang menjadi obyek penelitian. Informasi tersebut dapat diperoleh dari buku-buku, karya ilmiah dan sumber-sumber lain yang berkaitan dengan data mining dan algoritma C4.5

4. Hasil dan Pembahasan

4.1. Analisa Data Mining Untuk Prediksi Profit Perusahaan

Untuk mencapai tujuan akhir dari penelitian ini telah dilakukan pengumpulan data Proyek dan Estimasi Anggaran pengerjaan proyek. Bagian ini bertujuan untuk mengelompokkan data sehingga memudahkan dalam melakukan analisa berikutnya.

4.2. Implementasi Algoritma C4.5

Sistem yang akan dirancang dan digunakan dalam Prediksi Pencapaian Profit Pada Perusahaan. Data awal *numeric* maupun *nonnumeric* akan dibagi perkelas untuk memudahkan analisa berikutnya.

Setelah semua data yang akan dimasukkan dibagi perkelas, maka dilakukan proses klasifikasi dengan membuat pohon keputusan sebagai *output*.

Proses pengambilan keputusan untuk prediksi pencapaian profit pada perusahaan adalah sebagai berikut :

A. Biaya Langsung Personil

1. Biaya Tenaga Ahli
2. Biaya Tenaga Teknik
3. Biaya Tenaga Pendukung

B. Biaya Langsung Non Personil

1. Biaya Operasional Kantor
2. Biaya Peralatan
3. Biaya Pelaporan & Dokumen
4. Biaya Mobilisasi

Variabel yang akan menjadi keputusan adalah BAIK dan TIDAK BAIK.

Berikut adalah format data akhir yang akan di olah dengan algoritma C4.5 untuk memprediksi profit perusahaan tertera pada Tabel 1.

Pada Format data akhir pada tabel didapat berdasarkan dari *attribute* yang sudah dikelompokkan atau diklasifikasi, misalnya biaya tenaga ahli diklasifikasi kan menjadi 3 (tiga) bagian yaitu "*Budget*" apabila biaya dibawah dari anggaran, "*enough*" apabila biaya sesuai dengan anggaran, "*Past*", apabila biaya melewati anggaran yang telah ditetapkan begitupun dengan *attribute* lainnya.

4.3. Pohon Keputusan

Dari data pra-proses anggaran biaya, maka akan dilakukan klasifikasi data algoritma C4.5 dengan membuat pohon keputusan. Kasus yang tertera dalam tabel akan dibuat pohon keputusan untuk memprediksi *profit* perusahaan. Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus (1), sedangkan untuk menghitung nilai *entropy* dapat dilihat pada rumus (2).

Tabel 1. Data setelah pra-proses/format data akhir

Kode Proyek	Estimasi Anggaran							Tingkat Profit
	Biaya Lansung Personil			Biaya Langsung Non Personil				
	Tenaga Ahli	Tenaga Teknis	Tenaga Pendukung	OP Kantor	Peralatan	Laporan Dokumen	Mob	
A001	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Profit</i>
A002	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Defisit</i>
A003	<i>Past</i>	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Defisit</i>
A004	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Past</i>	<i>Defisit</i>
A005	<i>Enough</i>	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Profit</i>
A006	<i>Budget</i>	<i>Past</i>	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Defisit</i>
A007	<i>Enough</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Past</i>	<i>Past</i>	<i>Past</i>	<i>Defisit</i>
A008	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Profit</i>
A009	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Past</i>	<i>Defisit</i>
A010	<i>Enough</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Profit</i>
A011	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Past</i>	<i>Defisit</i>
A012	<i>Enough</i>	<i>Budget</i>	<i>Budget</i>	<i>Budget</i>	<i>Past</i>	<i>Budget</i>	<i>Budget</i>	<i>Profit</i>

Dalam pembuatan pohon keputusan, yang harus dilakukan adalah menghitung jumlah kasus, jumlah kasus untuk keputusan "Profit" (S1), jumlah kasus untuk keputusan "Defisit" (S2) dan kasus yang dibagi berdasarkan atribut biaya tenaga ahli, biaya tenaga teknik, biaya tenaga pendukung, biaya operasional kantor, biaya peralatan, biaya pelaporan dan dokumen, biaya peralatan, biaya mobilisasi. Setelah itu, lakukan perhitungan *gain* untuk setiap atribut. Adapun langkah-langkah pembuatan pohon keputusan adalah sebagai berikut :

1. 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 :

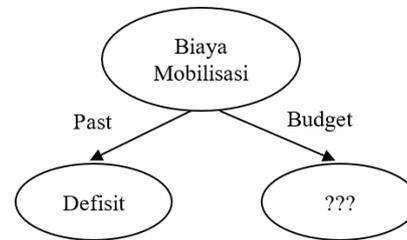
$$Entropy(total) = - \sum_{i=1}^n - P_i \times \log_2 P_i$$

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

*Entropy* (total) adalah menghitung nilai total keputusan *Profit* (5) dan *Defisit* (7), sedangkan 12 adalah jumlah keseluruhan kasus.

Setelah nilai *entropy* dan *gain* dihitung, kemudian hasil dari perhitungan tersebut dimasukkan ke dalam Tabel 2. Dari perhitungan Tabel 2 diketahui bahwa Biaya Mobilisasi merupakan *gain* terbesar dengan nilai 0.4764, maka Biaya Mobilisasi dijadikan sebuah akar, Pada Biaya Mobilisasi yang menyatakan Tidak

Baik semuanya terletak pada kolom Tidak Baik sedangkan pada kolom ya bernilai nol, maka dipastikan Keputusan tingkat profit dari Biaya Mobilisasi - Tidak Baik yaitu Tidak Baik, sedangkan pada Biaya Mobilisasi - Baik kolom Baik dan Tidak sama-sama berisi. Maka belum bisa kita pastikan Tingkat Profitnya dan masih perlu dilakukan perhitungan lagi, pohon keputusan dari tabel *node* 1 dapat dilihat pada Gambar 2.



Gambar 2. Pohon Keputusan Hasil Perhitungan *Node* 1

Selanjutnya adalah menyelesaikan untuk menghitung *Node* 1.1 sebagai akar, sama dengan cara yang di atas dengan menghitung nilai *entropy* dari atribut yang tersisa yaitu biaya tenaga ahli, biaya tenaga teknik, biaya tenaga pendukung, biaya operasional, biaya peralatan dan biaya pelaporan dan dokumen setelah dihitung *entropy*, kemudian menghitung *gain* untuk tiap-tiap atribut.

Setelah nilai *entropy* dan *gain* dihitung, kemudian hasil dari perhitungan tersebut dimasukkan ke dalam Tabel 3.

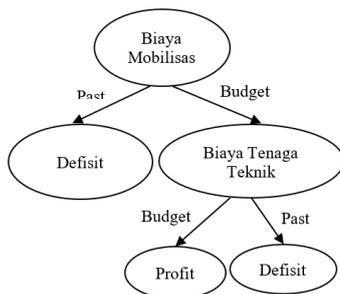
Tabel 2. *Perhitungan Node 1*

Node	Kasus	Profit	Defisit	Entropy	Gain
1 Total	12	5	7	0.9799	
B.Tenaga Ahli					
Budget	6	2	4	0.9183	0.2503
Enough	4	3	1	0.8113	
Past	2	0	2	0	
B. Tenaga Teknik					
Budget	7	4	3	0.9852	0.1043
Past	5	1	4	0.7219	
By.T Pendukung					
Budget	9	4	5	0.9911	
Past	3	0	3	0	0.2366
By.OP.Kantor					
Budget	9	4	5	0.9911	0.0070
Past	3	1	2	0.9183	
By.Peralatan					
Budget	8	4	4	1	0.0428
Past	4	1	3	0.8113	
By.Pelaporan					
Budget	9	5	4	0.9917	0.2361
Past	3	0	3	0	
By.Mobilisasi					
Budget	7	5	2	0.8631	0.4764
Past	5	0	5	0	

Tabel 3. Perhitungan Node 1.1

Node	Kasus	Profit	Defisit	Entropy	Gain
1.1 Total	7	5	2	0.8631	
B.Tenaga Ahli					
Budget	3	2	1	0.9183	0.4696
Enough	3	3	0	0	
Past	1	0	1	0	
B. Tenaga Teknik					
Budget	4	4	0	0	0.4697
Past	3	1	2	0.9183	
By.T Pendukung					
Budget	6	5	1	0.6499	0.3060
Past	1	0	1	0	
By.OP.Kantor					
Budget	6	4	2	0.9179	0.0763
Past	1	1	0	0	
By.Peralatan					
Budget	5	4	4	0.5150	0.3523
Past	2	1	2	0.5	
By.Pelaporan					
Budget	6	5	1	0.6500	0.3059
Past	1	0	1	0	

Dari hasil Tabel 3 dapat diketahui bahwa atribut *gain* tertinggi adalah Biaya Tenaga Teknik dengan nilai 0.4697 berarti Biaya Tenaga Teknik menjadi akar selanjutnya. nilai atribut Budget adalah *Profit* dan atribut *Past* adalah *Defisit*, karena nilai atribut biaya tenaga teknik telah diketahui jadi pada node 1.1 sudah ditemukan pohon keputusan akhir, Pohon keputusan yang terbentuk dapat dilihat pada Gambar 3 berikut ini:



Gambar 3. Pohon Keputusan Hasil Perhitungan Node 1.1

Adapun aturan atau *rule* yang terbentuk berdasarkan pohon keputusan terakhir seperti pada gambar 3 di atas adalah sebagai berikut:

1. Jika Biaya Mobilisasi = melebihi anggaran, Maka Perusahaan akan mengalami kerugian (*defisit*)
2. Jika Biaya Mobilisasi = kurang atau sama dengan anggaran dan Biaya Teknik = melebihi anggaran, Maka Perusahaan akan mengalami kerugian (*Defisit*)
3. Jika Biaya Mobilisasi = kurang dari anggaran, Biaya Tenaga Teknik = kurang atau sama dengan anggaran, Maka Perusahaan akan mendapatkan Keuntungan (*Profit*).

Berdasarkan dari *rule/knowledge* yang dihasilkan terdapat beberapa *rules* cukup sesuai dengan kejadian yang terjadi di antaranya:

1. Hasil dari pengolahan metode *Data Mining* dengan *algoritma C4.5* untuk Prediksi Profit Pada Perusahaan yaitu PT SMOE Indonesia di karenakan oleh biaya Mobilisasi dan biaya tenaga teknik yang sering melebihi dari anggaran yang telah di tetapkan sehingga perusahaan mengalami kerugian pada akhir pelaksanaan proyek.
2. Pada Estimasi Anggaran biaya PT SMOE Indonesia pada tahun 2017 jika pelaksanaan proyek menekan biaya mobilisasi dan biaya tenaga teknik dengan baik tanpa merugikan pekerjaan yang dilaksanakan maka perusahaan akan mendapatkan keuntungan (*profit*).

Melalui proses data mining yang telah dilakukan menggunakan Algoritma C.45 dengan *software Data Mining open source WEKA* diketahui informasi sebagai pengetahuan baru berupa 6 *rule*/aturan dengan tingkat kinerja algoritma C.45 sebesar 92,60% +/- 1,36%. *Rule* yang ditemukan yang dapat dijadikan landasan bagi pengelola perusahaan dalam melakukan prediksi pencapaian target profit sehingga pengelola perusahaan dapat mengantisipasi dengan mengambil langkah-langkah bisnis yang tepat dalam mengupayakan tercapainya profit.

### 5. Kesimpulan

Berdasarkan pembahasan yang telah dilakukan dengan teknik data mining dengan menggunakan algoritma C4.5 maka peneliti menarik beberapa kesimpulan yang penting. Adapun kesimpulan-kesimpulan tersebut adalah sebagai berikut: (1). Algoritma C4.5 dapat digunakan sebagai metode

klasifikasi dalam memprediksi Profit Perusahaan PT SMOE Indonesia dengan memperhatikan nilai *gain* (penguatan) tertinggi dari tujuh atribut yaitu Biaya Tenaga Ahli, Biaya Tenaga Teknik, Biaya Tenaga Pendukung, Biaya Operasional Kantor, Biaya Peralatan, Biaya Pelaporan dan Dokumen, dan Biaya Mobilisasi; (2). Pohon Keputusan yang dihasilkan oleh teknik klasifikasi Algoritma C4.5 dari variabel yang memiliki gain tertinggi yaitu Biaya Mobilisasi dan Biaya Tenaga Teknik, sehingga perusahaan dapat menjadikan acuan dalam pengaturan dan penggunaan dana anggaran biaya untuk proyek yang akan datang.

#### Daftar Pustaka

- Faradillah, S., 2013. Implementasi data mining untuk pengenalan karakteristik transaksi customer dengan algoritma C4.5. *Pelita Informatika Budi Darma*, 63–70.
- Gamarrá, C., Guerrero, J. M., Montero, E., 2016. A knowledge discovery in databases approach for industrial microgrid planning. *Renewable and Sustainable Energy Reviews*, 60, 615–630.
- Gunadi, G., Sensuse, D.I., 2012. Penerapan metode data mining market basket analysis terhadap data penjualan produk buku dengan menggunakan algoritma apriori dan frequent pattern growth FP-GROWTH), 4(1)
- Haryati, S., Sudarsono, A., Suryana, E., 2015. Implementasi data mining untuk memprediksi masa studi mahasiswa menggunakan algoritma C4.5 (studi kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama Vol., 11(2)*, 130–138.
- Lestari, S., Suryadi, A., 2014. Model klasifikasi kinerja dan seleksi dosen berprestasi dengan klasifikasi algoritma C.45. *Prosiding Seminar Bisnis & Teknologi*, 15–16.
- Santoso, H., Hariyadi, I. P., Prayitno, 2016. Data mining analisa pola pembelian produk. *Teknik Informatika*, (1), 19–24.
- Wowor, A. S., Mangantar, M., 2012. Laba bersih dan tingkat risiko harga saham pengaruhnya terhadap Dividen pada perusahaan otomotif yang terdaftar Di bursa efek indonesia, 2(4), 13–23. <https://doi.org/10.1002/eji.201444988>
- Zulkifli, A., 2016. Metode C45 Untuk Mengklarifikasi Pelanggan Perusahaan Telekomunikasi Seluler, 2(1), 65–76.