

**Penerapan Algoritma C4.5 Pada Analisis Penentuan Kelayakan Penerima Bonus Tahunan
Pegawai Bank Sinarmas Multifinance**

Baiq Andriska Candra Permana¹, Intan Komala Dewi Patwari²

Fakultas Teknik Universitas Hamzanwadi
andriska.cp@gmail.com¹, intankdp19@gmail.com²

Abstrak

Data mining merupakan suatu rangkaian proses yang memberikan nilai tambah berupa informasi terhadap suatu kumpulan data yang selama ini tidak diketahui secara manual. Pada data mining dilakukan penggalian data dengan tujuan memanipulasi data tersebut agar menjadi informasi yang lebih berharga, yang dikemudian hari diharapkan dapat bermanfaat untuk membantu memutuskan kelayakan pegawai dalam menerima bonus disetiap tahunnya. Kelayakan pegawai untuk menerima bonus setiap tahun menggunakan algoritma C4.5 dilihat berdasarkan beberapa variable atau kriteria yang sudah ditetapkan dan akan menghasilkan suatu pohon keputusan. Dari pohon keputusan tersebut akan diambil pengetahuan baru berupa aturan yang dijadikan sebagai pola dalam pengambilan keputusan. Teknik seperti ini nantinya akan dapat membantu seorang pemimpin dalam mengambil suatu keputusan apakah seorang pegawai layak atau tidak untuk diberi bonus.

Kata Kunci : C4.5 , Data mining, Klasifikasi

Abstract

Data mining is a series of processes that add a new value in the form of information to a data set that has not been known manually. Data mining aims to manipulate the data so that it becomes more valuable information, which in the future is expected to be useful to help decide the employee's eligibility in receiving bonuses every year. The feasibility of employees to receive bonuses every year is determined using the C4.5 algorithm based on several variables or criteria that have been established and will produce a decision tree. From the decision tree, new knowledge in the form of rules will be taken as a pattern in decision making. This Techniques will able to help a leader to make a decision whether an employee is worthy or not to be given a bonus.

Keyword : C4.5 , Data mining, Classification

1. Pendahuluan

Dalam upaya peningkatan kualitas serta loyalitas pegawai yang dimiliki, Bank Sinarmas Multifinance di Lombok Timur sekali dalam setahun memberi bonus kepada pegawai yang memenuhi kriteria yang telah ditentukan oleh pimpinan. Karena banyaknya pegawai dan kriteria yang harus di analisa, pemimpin mengalami kesulitan untuk memberikan keputusan terhadap pegawai yang layak atau tidak untuk menerima bonus tersebut. Oleh sebab itu perlu adanya pengolahan data yang memberikan pola keputusan yang dapat membantu pemimpin dalam memberikan keputusan.

Algoritma C4.5 merupakan salah satu solusi yang dapat digunakan untuk memecahkan masalah terkait kasus pada teknik klasifikasi. Keluaran yang diberikan algoritma C4.5 berupa suatu pohon keputusan yang terstruktur, yang dapat digunakan dalam mengubah kumpulan data menjadi suatu pohon keputusan yang terdiri dari suatu aturan – aturan keputusan.

2. Tinjauan Pustaka

2.1 Penelitian Terkait

Penelitian dilakukan oleh Nurul Azwanti (2018), pada penelitian yang dilakukan ditujukan untuk memprediksi mahasiswa yang mengulang pada mata kuliah di AMIK Labuhan Batu dengan menggunakan algoritma C4.5. Pengujian dilakukan dengan WEKA memberi keluaran yang menunjukkan bahwa hasil lebih efektif dan

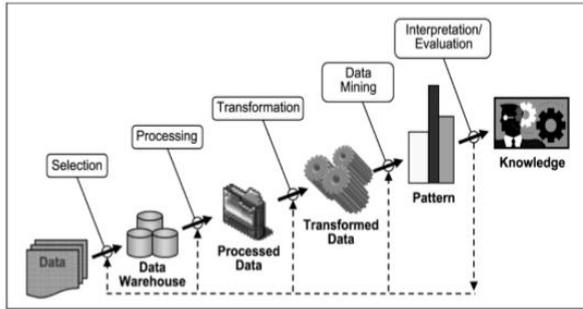
fleksibel karena pohon keputusan yang terbentuk hampir sama dengan perhitungan manual dan hasil sangat membantu dalam melakukan keputusan.[1]

Penelitian dilakukan Winda, Yulia dan Wahyu (2016) melakukan penelitian untuk penentuan jalur kelulusan mahasiswa. Pada penelitian ini dilakukan pembuatan model menggunakan data mahasiswa tingkat akhir, diperoleh pohon keputusan yang terdiri dari 7 rule aturan yaitu 3 aturan untuk jalur skripsi dan 4 aturan untuk non skripsi. Akurasi di peroleh 0,85 – 1,00 sehingga dapat dikatakan metode ini dapat diterapkan dalam penentuan jalur kelulusan mahasiswa tingkat akhir.[2]

2.2 Landasan Teori

1. Data Mining

Data mining merupakan suatu proses analisa data untuk menemukan suatu pola dari kumpulan data tersebut. Data mining mampu menganalisa data yang besar menjadi informasi berupa pola yang mempunyai makna bagi pendukung keputusan.(rika). Data mining disebut juga sebagai kegiatan pengumpulan, pemakaian data historis untuk menemukan keteraturan pola atau hubungan dalam data yang berukuran besar. Keluaran yang diberikan dapat digunakan sebagai acuan untuk membuat suatu keputusan. Tahapan data mining ditunjukkan oleh gambar berikut [3] :



Gambar 1. Proses pada data mining

2. Algoritma C4.5

Algoritma C4.5 atau yang biasa di sebut sebagai decision tree memiliki training sample berupa sekumpulan data yang nantinya akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sample merupakan field – field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data [4].

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

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk setiap nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cangan memiliki kelas yang sama.

3. Klasifikasi

Klasifikasi merupakan suatu proses mencari suatu himpunan model (fungsi) yang dapat mendeskripsikan dan membedakan kelas-kelas data atau konsep – konsep dengan tujuan dapat menggunakan model tersebut untuk melakukan prediksi kelas dari suatu objek yang mana kelasnya belum diketahui.

4. Entropy Dan Information Gain

Suatu objek yang diklasifikasikan dalam sebuah pohon harus dites nilai Entropynya. Entropy merupakan ukuran dari terori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* kumpulan data. Dari nilai Entropy tersebut di hitung nilai *Information gain*. Pemilihan atribut sebagai simpul, baik akar atau simpul internal didasarkan pada nilai *Gain* tertinggi dari atribut yang ada. Rumus menghitung nilai entropy menggunakan persamaan [5]

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

Keterangan :

S = himpunan kasus

n = jumlah partisi atribut A

Pi = proporsi Si terhadap S

|Si| = jumlah kasus pada partisi ke i

|S| = jumlah kasus dalam S

A= atribut

Rumus untuk mencari nilai gain :

$$Gaint(S,A) = \sum_{f=1}^n \frac{|S_i|}{|S|} + Entropy(S_i) \dots\dots(2)$$

3. Metode Penelitian

3.1 Metode Pengumpulan Data Awal

Dalam penyusunan penelitian ini metode yang digunakan dalam proses pengumpulan data antara lain :

- a. Studi Pustaka

Studi pustaka dilakukan dengan melakukan pengumpulan informasi yang sesuai dengan topik atau masalah yang sedang diteliti.

Informasi didapatkan melalui buku ilmiah, jurnal penelitian, tesis dan disertasi serta sumber – sumber tercetak maupun elektronik.

b. Dokumentasi

Pengumpulan data dilakukan dengan cara pengumpulan dokumen tertulis maupun dokumen elektronik dari lembaga / institusi. Dokumen ini diperlukan guna mendukung kelengkapan penelitian.

c. Wawancara

Wawancara dilakukan secara langsung dengan kepala cabang sebagai pihak yang terkait secara langsung dengan penelitian yang sedang dilakukan dengan cara mengajukan pertanyaan sesuai kebutuhan. Informan dapat menyampaikan informasi sebagaimana yang diharapkan oleh peneliti.

3.2 Seleksi Data

Seleksi data merupakan pemilihan data yang akan digunakan dalam proses algoritma klasifikasi C4.5. Tujuan dari seleksi data adalah mendapatkan himpunan data target, pemilihan himpunan data atau memfokuskan data pada subset variable atau sample data dimana penemuan akan dilakukan.

3.3 Transformasi Data

Transformasi data adalah proses mentransformasikan atau mengubah data ke dalam bentuk yang sesuai agar dapat di proses dengan algoritma C4.5

3.4 Penentuan Information Gain

Perhitungan semua atribut/variable untuk mendapatkan informasi gain tertinggi yang nantinya akan dijadikan sebagai akar pada pembuatan pohon keputusan.

3.5 Pohon Keputusan

Pohon keputusan adalah hasil dari proses perhitungan entropy dan information gain , setelah perhitungan dilakukan berulang ulang sampai semua atribut pohon memiliki kelas dan tidak bisa lagi dilakukan proses perhitungan.

3.6 Aturan Rule /Model

Aturan – aturan / rule model dalam uraian penjelasan yang merepresentasikan sebuah pohon keputusan.

4. Hasil Dan Pembahasan

4.1 Pengumpulan Data Awal

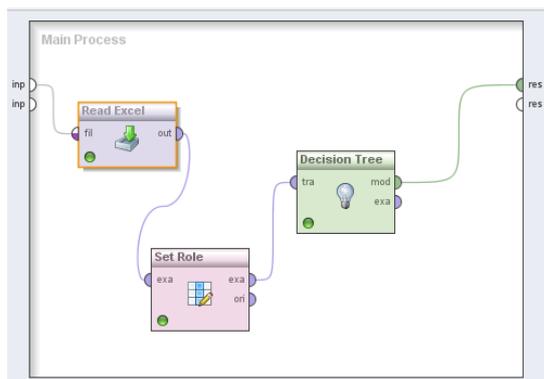
Pada tahap awal, data di peroleh dari pengumpulan data yang didapatkan dari Bank Sinarmas Multifinance. Data yang di peroleh sebanyak 130 sample data. Dari data tersebut , selama ini kriteria yang dijadikan sebagai acuan pimpinan untuk memutuskan seorang kariawan berhak untuk mendapatkan bonus tahunan adalah berdasarkan 4 variable. Keempat variable tersebut adalah : masa kerja, kehadiran, loyalitas, dan hasil kerja. Berikut adalah beberapa gambaran dari sample yang digunakan dalam penelitian ini :

No	Nama	masa kerja (bln)	Kehadiran	loyalitas	Hasil pekerjaan	ket
1	Abidatus Solehah	36	80%	75%	90	layak
2	Akbar Hudatulloh	12	80%	65%	90	tidak
3	Alfiana	5	80%	75%	90	tidak
4	alwan holis	9	75%	80%	90	tidak
5	andika erlangga	36	80%	70%	90	layak
6	Andri Saputra	24	70%	80%	80	layak
7	Anisa Zuhriyah	11	65%	65%	70	tidak
8	ardianto	4	77%	80%	70	tidak
9	ari susanto	12	85%	75%	70	layak
10	Asri	12	85%	65%	65	tidak
11	Astuti	5	85%	65%	65	tidak
12	Budi alamsyah	4	80%	70%	90	tidak
13	Budiawan Saputra	11	70%	80%	90	tidak
14	Dewi anggraini	36	85%	70%	75	layak
15	Dian safitri	36	70%	80%	75	layak

Gambar 2. sample data set penerimaan bonus tahunan pegawai Bank Sinarmas Multifinance

4.2 Pengolahan Dengan C4.5

Pengolahan data menggunakan algoritma C4.5 menggunakan tools rapidminer ditunjukkan oleh gambar berikut :

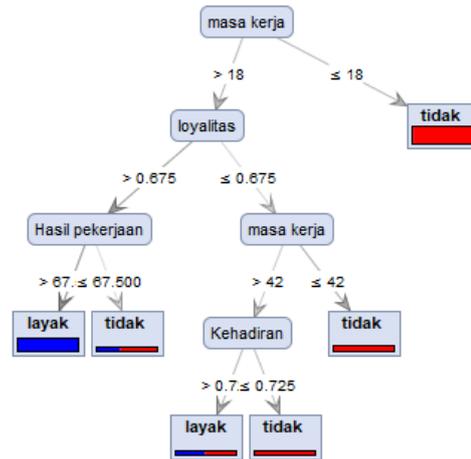


Gambar 3. Pengolahan data menggunakan algoritma Algoritma C4.5

Diperoleh pohon keputusan sebagaimana yang ditunjukkan pada gambar dengan pola – pola tertentu. Pada gambar ditunjukkan bahwa apabila masa kerja kurang dari 18 bulan maka pegawai tidak layak mendapatkan bonus.

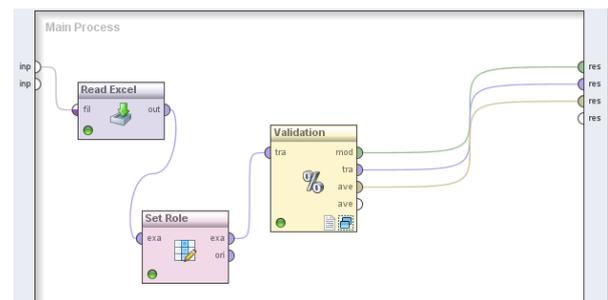
- a. Jika loyalitas > 67% dan hasil kerja >67% maka layak
- b. Jika loyalitas > 67% dan hasil kerja <=67% maka tidak layak
- c. Jika loyalitas <= 67,5 % dan masa kerja >42 bulan dan kehadiran >70% maka layak

- d. Jika loyalitas <= 67,5 % dan masa kerja >42 bulan dan kehadiran <=72,5% maka tidak layak
- e. Jika loyalitas <=67,5% dan masa kerja <=42 bulan maka tidak layak.

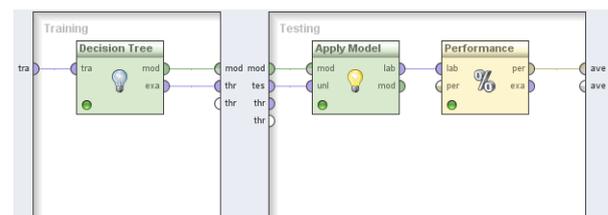


Gambar 4. Pohon keputusan kelayakan penerima bonus tahunan

Tahap selanjutnya adalah menentukan nilai akurasi dari algoritma C4.5 dalam menentukan kelayakan pegawai dalam menerima bonus sebagaimana ditunjukkan gambar:



Gambar 5 Proses Utama penentuan akurasi



Gambar 6 Training data menggunakan algoritma C4.5

accuracy: 94.62% +/- 4.93% (mikro: 94.62%)			
	true layak	true tidak	class precision
pred. layak	44	2	95.65%
pred. tidak	5	79	94.05%
class recall	89.80%	97.53%	

Gambar 7. Tabel akurasi

4.3 Evaluasi Dan Validasi Hasil

Berdasarkan eksperimen yang telah dilakukan diperoleh nilai akurasi untuk penggunaan algoritma C4.5 adalah 94,62% dimana pada tabel akurasi didapatkan bahwa dari 44 prediksi layak ternyata mendapatkan data layak dan 2 data error yang di prediksi layak namun tidak layak. Sementara dari 79 data yang di prediksi tidak layak terdapat 5 error yang ternyata layak. Selain itu di dapatkan kesimpulan bahwa dari sekian variable penentu kelayakan penerima bonus, masa kerja merupakan faktor penentu utama layak tidaknya seorang pegawai mendapat bonus tahunan.

5. Kesimpulan

Penggunaan algoritma C4.5 untuk memutuskan pegawai yang berhak untuk mendapatkan bonus tahunan berhasil memberikan akurasi sebesar 94,62%. Hal ini membuktikan bahwa penggunaan algoritma ini dapat membantu pimpinan untuk memberikan keputusan. Selain itu dari analisa yang dilakukan diperoleh suatu fakta bahwa dari beberapa kriteria yang di gunakan untuk memberikan keputusan, lama masa kerja merupakan faktor utama sebagai penentu layak atau tidaknya seorang pegawai mendapatkan bonus tahunan. Untuk penilaian

selanjutnya, sebaiknya sample data ditingkatkan untuk mendapatkan akurasi yang lebih baik.

6. Daftar Pustaka

- [1] C. Algoritma *et al.*, "Algoritma c4.5 untuk memprediksi mahasiswa yang mengulang mata kuliah (studi kasus di amik labuhan batu)," 2018.
- [2] W. W. Ariestya *et al.*, "Decision Tree Learning Untuk Penentuan Jalur Kelulusan Mahasiswa," 2010.
- [3] J. T. Informatika-s and F. I. Komputer, "Clusterisasi Data Pasien Menggunakan K-Means Clustering Tri Hertanto," no. 5.
- [4] E. P. Cynthia, "MENGKLASIFIKASI DATA PENJUALAN BISNIS GERAJ," no. July, 2018.
- [5] G. Syahputra, M. Kom, P. Studi, and M. Informatika, "Penerapan Algoritma C4 . 5 Dalam Analisa Kelayakan Penerima Bonus Tahunan Pegawai (Studi Kasus : PT . Multi Pratama Nauli Medan)," 2015.