

PENGAMBILAN KEPUTUSAN PEGAWAI TIDAK TETAP MENJADI PEGAWAI TETAP DENGAN *DECISION TREE*

Febryantahanuji¹, Irwan Sembiring² dan Hindriyanto Dwi Purnomo³

Universitas Kristen Satya Wacana Salatiga

Jl. Diponegoro No.52-60 Salatiga, Jawa Tengah

E-mail : febryan@stekom.ac.id¹, irwan@staff.uksw.edu², hindriyanto.purnomo@staff.uksw.edu³

ABSTRAK

Rekrutmen karyawan merupakan hal yang sangat sensitif di dalam sebuah perusahaan, oleh karena itu setiap perusahaan yang melakukan rekrutmen karyawan harus memperhatikan betul faktor yang mempengaruhi kelayakan seorang karyawan untuk dijadikan pegawai tetap, agar tidak terjadi kesalahan dalam masalah ini perlu adanya sebuah sistem untuk meminimalisir pengangkatan pegawai outsourcing menjadi pegawai tetap disebuah perusahaan. Penerapan metode *decision tree C4.5* terhadap seleksi penerimaan pegawai tetap dapat membantu sebuah perusahaan untuk menentukan pegawai yang layak dijadikan pegawai tetap dengan memprediksi maupun peramalan data dengan membuat struktur pohon yang memiliki kode dari hasil uji atribut, dengan menggunakan algoritma C4.5 didapatkan indikator utama dalam penentuan karyawan tetap.

Kata Kunci: Rekrutmen karyawan, *decision tree*, c4.5

ABSTRACT

Employee recruitment is something sensitive in a company. Therefore, the company which hiring an employee have to notice the factor which affect eligibility that employee to be used as a permanent employee. So, to ignore an error, a company should have a system to help them to minimalize recruitment from outsourcing employee to be a permanent employee in a company. The application *decision tree C4.5* method in a employee recruitment selection can help a company for determining an employee which is eligibility with predicted as well as portend data with make tree structure which have code from result of that attribute test, using algoritma C4.5 it can get a main indicator in a employee recruitment.

Keyword: *Employee recruitment*, *decision tree*, c4.5

PENDAHULUAN

Dalam kompetisi perusahaan yang kian besar, dimana antar kompetitor saling mengambil karyawan, adalah hal yang tidak bisa dielakkan. Sebab, di perusahaan yang berkaitan dengan jasa dan kreativitas, karyawan bertalenta adalah aset sehingga sering disebut sebagai human capital, dengan demikian perusahaan menginginkan ketika penerimaan karyawan mendapatkan sebuah karyawan yang bertalenta menjadi salah satu tantangan Profesional SDM, (Hamidah Jantan, 2010). Karena tugas ini melibatkan banyak manajerial untuk memutuskan karyawan yang tepat bagi perusahaan, dengan keputusan tersebut

sangatlah tidak mudah, karena perlu pertimbangan dari berbagai indikator-indikator seperti kinerja, pengalaman kerja, loyalitas, kedisiplinan dan tingkat pendidikan dengan demikian penulis ingin menerapkan sebuah teknik data mining untuk membantu bagian penerimaan karyawan, dengan beberapa metode yang pernah dipakai untuk menyelesaikan masalah penerimaan karyawan yaitu: Id3, CART, C4.5. dari hasil study literatur, dengan metode C4.5 diharapkan dapat digunakan untuk teknik pengambilan keputusan dimana hasil dari metode berbentuk *tree* sehingga dapat diketahui indikator prioritas dalam penentuan penerimaan karyawan. Penerapan

Algoritma dalam sistem informasi sendiri sudah tidak asing lagi di dunia bisnis termasuk pemerintah, dalam penelitian Sembiring (2004) menyatakan bahwa dalam membantu keputusan kelayakan investasi *e-government* menggunakan metode ROI (*Return On Investment*) dengan menambahkan dua domain penting yaitu *bussiness domain* dan *technology domain*, dapat disimpulkan bahwa layanan masing-masing *e-government* layak atau tidak (Sembiring, 2004)

Dalam permasalahan yang terjadi pada pengangkatan karyawan tetap di perusahaan studi kasus adalah tidak adanya prestasi atau talenta yang sesuai keinginan perusahaan, dikarenakan dalam rekrutmen pegawai masih terjadi kecurangan yaitu adanya tindakan nepotisme yang dilakukan oleh beberapa oknum yang tidak bertanggung jawab. Dimana dampak dari tindakan tersebut dapat membuat semangat kerja pegawai menurun dan menimbulkan ketidakadilan antar pegawai atau karyawan, dimana pegawai yang mempunyai prestasi baik belum tentu akan terpilih menjadi pegawai tetap dikarenakan tidak mempunyai kerabat atau kedekatan dengan atasan. Penerapan pohon keputusan juga pernah dilakukan dalam pengambilan calon nasabah untuk penentuan kredit, dimana terdapat sebuah hasil pengambil kebijakan dalam menentukan nasabah yang layak mendapatkan kredit dan nasabah yang tidak layak mendapatkan kredit (Cahyo Dimas K, 2017). Dengan adanya permasalahan yang ada, penulis membuat sebuah sistem pendukung keputusan dengan 5 indikator yang saling mempengaruhi, yaitu kinerja, pengalaman kerja, loyalitas, kedisiplinan, dan tingkat pendidikan. Indikator tersebut yang akan digunakan untuk merancang sebuah sistem pendukung keputusan pengangkatan pegawai tidak tetap menjadi pegawai tetap dengan metode *decission tree c4.5* pada sebuah perusahaan agency tenaga kerja. Dengan adanya sistem pendukung keputusan ini diharapkan akan meminimalisir kecurangan yang dilakukan

oleh oknum nakal, agar pegawai dapat meningkatkan prestasi kerjanya dan perusahaan akan mendapatkan pegawai yang berkualitas karena pegawai yang mempunyai prestasi baik yang akan terpilih sebagai pegawai tetap.

TINJAUAN PUSTAKA

2.1. Kajian Penelitian Yang Relevan

Metode C4.5 sudah diterapkan dalam beberapa penelitian, antara lain H. Jantan, dkk (2010) juga pernah melakukan penelitian bahwa menggunakan metode C4.5 *Classification Algorithm* sebagai Penentuan Klasifikasi bakat atau memprediksi bakat manusia, sehingga dapat disimpulkan dengan metode tersebut didapatkan sebuah teknik dalam pengambilan keputusan yang tepat. Penggunaan metode C4.5 juga diperkuat dalam penelitian N. Sivaram, K. Ramar (2010) dengan metode yang digunakan Fuzzy *C-means*, *K-Means*, *Id3*, *C4.5*, yang digunakan untuk mengatasi dalam Pengelompokan data dengan *fuzzy C-means clustering* dan *Algorithm K-means clustering* dan klasifikasi menggunakan algoritma *decission tree*, menganalisis masalah yang terlibat dalam proses rekrutmen lulusan baru, dan mencari cara untuk menghemat waktu dan biaya. dapat disimpulkan bahwa Algoritma C4.5 memiliki akurasi yang lebih baik. H. Jantan, dkk (2011) menerapkan metode C4.5 untuk menyelesaikan sebuah permasalahan dalam memprediksi kinerja karyawan, penelitian tersebut memberikan kesimpulan bahwa C4.5 *decission tree* memberikan nilai akurasi yang paling tinggi dibandingkan dengan metode yang ada yaitu metode C4.5 *decission tree*, *Random Forest*, *Multilayer Perceptron (MLP)* and *Radial Basic Function Network*. Qasem A. Al-Radaideh, Eman Al Nagi (2012) dengan metode *ID3*, *Naïve Bayes*, C4.5 akan dibandingkan dan digunakan untuk membantu dalam pengambilan keputusan serta penentuan indikator berpengaruh dalam mengevaluasi kinerja karyawan, dari hasil pengujian

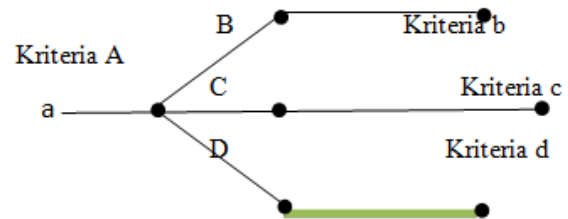
dapat disimpulkan C4.5 *decision tree* memiliki akurasi yang paling baik diantara metode yang telah dibandingkan. Adapun penelitian menurut John M. Kirimi, Christopher A. Moturi (2016) juga mencoba mencoba untuk mengatasi masalah yang terjadi dalam Penentuan indikator yang berpegaruh pada kinerja karyawan, dengan indikator umur, kulaifikasi, jenis kelamin, status perkawinan, pelatihan dan skor kinerja dapat disimpulkan, dari algoritma yang dipakai (ID3, *Naïve Bayes*) peneliti menyatakan bahwa C4.5 *decission tree* memiliki nilai akurasi paling tinggi. Dari kajian penelitian relevan yang sudah pernah dilakukan, terdapat perbedaan kasus dan indikator penentu keputusan terhadap penelitian yang akan dilakukkan oleh penulis, dimana pada penelitian terdahulu tidak adanya indikator kinerja, pengalaman kerja, loyalitas, kedisiplinan, dan tingkat pendidikan dalam mengambil keputusan untuk menentukan. pegawai tidak tetap menjadi pegawai tetap disebuah perusahaan agency tenaga kerja.

2.2. Decision Tree

Decision Tree merupakan salah satu teknik data mining untuk memprediksi maupun peramalan data, metode tersebut digunakan oleh penulis dikarenakan *Decision tree* dapat membuat struktur pohon yang memiliki kode dari hasil uji atribut, dimana setiap cabang merupakan pernyataan yang harus dipenuhi dari tiap ujung pohon (Kelas Data). *Node* teratas dari *Decission tree* disebut sebagai *root* (Akar Pohon) yang akan digunakan sebagai indikator penentu percabangan dan memiliki pengaruh paling besar di kelas data.

Umumnya *Decission Tree* dalam melakukan proses data secara *Top-Down* sebagai solusi dengan cara melakukan klasifikasi data baru, yang belum diketahui label dan akan di uji dengan cara menentukan jalur *root* sampai node akhir sehingga dapat menghasilkan sebuah

prediksi label dari data baru tertentu (Julianto, Yinitarini, & Sophan, 2014).

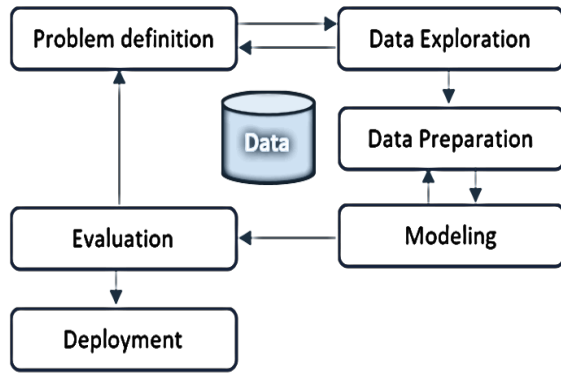


Gambar 1. Susunan Pohon Keputusan (Julianto, Yinitarini, & Sophan, 2014)

2.3. Data Mining.

Data mining merupakan salah satu bidang paling penting dalam penelitian yang bertujuan untuk memperoleh informasi dari data *set*, dimana sekumpulan data yang besar yang di ekstrak sehingga menghasilkan sebuah informasi baru yang dapat dimanfaatkan sesuai kebutuhan. Data mining mulai ada sejak 1990-an sebagai cara yang efektif untuk mengambil pola dan informasi yang sebelumnya tidak diketahui dari suatu data *set*. Teknik data mining digunakan untuk menemukan hubungan antara data untuk melakukan pengklasifikasi nilai-nilai dari beberapa variabel, membagi data yang diketahui menjadi kelas-kelas data yang mempunyai kesamaan karakteristik (*Clasification*). Data mining merupakan bagian dari proses penemuan pengetahuan dari basis data (*Knowledge Discovery in Databases*), yang mana tahapan dari *Knowledge Discovery in Databases*.

Dalam penerapan data mining memiliki parameter untuk data *test* diantaranya asosiasi, klasifikasi dan *clustering*. Data mining melibatkan langkah kunci yang meliputi *problem definition, data exploration, data preparation, modeling, serta evaluating and deployment*. (Mustofa & Suasana, 2018).



Gambar 2. Proses Data Mining (Mustofa & Suasana, 2018).

Keterangan:

Data : Data merupakan sebuah kumpulan informasi yang akan digunakan untuk penentuan indikator, yang terdapat atribut, dan *record* data karyawan.

Problem definition : Menganalisa sebuah masalah yang terjadi dengan melihat data Training Karyawan atau sumber data.

Data Exploration : Melakukan perubahan data yang akan digunakan untuk sebuah tahap proses pengolahan, disini penulis menggunakan format .xls

Data Preparation : Persiapan data yang akan digunakan dalam proses sebuah penerapan metode, apakah data yang digunakan sudah sesuai dengan penggunaan metode atau belum.

Evaluation : Tahapan ini digunakan untuk pengujian sebuah

metode yang akan digunakan, dimana penulis menggunakan *X-Fold Cross Validation* untuk mendapatkan sebuah *accuracy* dan *performance* dari sebuah pengujian

Deployment : Penerapan dari hasil pengolahan data training untuk diterapkan pada data baru.

2.4. *Algoritm C.45*

C.45 Merupakan turunan dari metode *clasificasion Tree* yang digunakan untuk membuat klasifikasi data yang bersifat prediktif, C4.5 salah satu algoritma *mechine learning* Algoritma C4.S merupakan salah satu algoritma *machine learning* (Pandiangan, 2017). Dengan algoritma ini akan memberikan sekelompok data yang tidak memiliki label (*Unsupervised Learning*) yang akan di proses dan ditentukan label sehingga menghasilkan data training, kemudian hasil dari proses data akan digunakan untuk mengolah data yang baru (data testing), dalam pengklasifikasian algoritma C4.5 data testing yang di proses akan menghasilkan sebuah kelas sebagai data *prediction*. Langkah-langkah dalam penyelesaian algoritma C4.5:

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

$$Entropi(S) = \sum_{j=1}^k -p_j \log_2 p_j \dots \dots \dots (1)$$

Keterangan :

S : Himpunan Kasus

K : Jumlah Partisi Kasus.

P_j : Probabilitas yang didapat dari jumlah (Ya/Tidak) dibagi total kasus.

Mencari gain.

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

Keterangan :

- S : Himpunan kasus
- A : Atribut
- N : j umlah partisi atribut A
- |S_i| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

ANALISIS DAN HASIL PENGUJIAN

3.1 Pengumpulan data

Data yang digunakan adalah data kuantitatif yang berasal dari perusahaan agency tenaga kerja, adapun teknik data maining salah satunya adalah metode klasifikasi yang meliputi .algoritma C4.5, Support Vector Machines (SVM), AdaBoost, k-Nearest Neighbor (k-NN), Naive Bayes dan CART. Metode klasifikasi memproses sebuah data dengan meletakkan data kedalam kelas yang sudah didefinisikan sebelumnya. Dari beberapa metode klasifikasi dalam pembuatan *decission tree* yang ada seperti ID3, C5.0, dan CART, pada masalah klasifikasi terdapat 2 jenis variabel, yaitu target variable (yang menunjukkan sebuah objek masuk dalam kategori yang mana), dan variabel prediktor (yaitu data para objek yang akan dijadikan dasar untuk memperkirakan objek tersebut masuk kelompok mana dan C4.5 merupakan algoritma yang dapat membuat cabang *tree* sebanyak 2 atau lebih, tergantung kandidat list yang dirancang. agar mendapatkan pohon keputusan yang merupakan solusi dari sebuah sistem yang manusia kembangkan dalam pencarian keputusan untuk mengatasi masalah-masalah tersebut serta memperhitungkan berbagai macam faktor yang ada di dalam lingkup permasalahan.

Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah. Peranan pohon keputusan sebagai alat bantu dalam mengambil keputusan (*decission support tools*) telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori graf. Kegunaan pohon keputusan yang sangat banyak ini membuatnya telah dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan.

Untuk mendapatkan keakuratan prediksi dengan metode C4.5, algoritma tersebut memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik serta dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model.

Data pegawai tidak tetap dengan indikator (kinerja, pengalaman kerja, loyalitas, kedisiplinan dan tingkat pendidikan) tersebut akan digunakan sebagai data *training* untuk pembuatan *decission tree* yang akan digunakan untuk memprediksi penerimaan pegawai tetap dan mengetes tingkat keberhasilan, prediksi tersebut menggunakan data testing terhadap pegawai yang memiliki pengalaman pekerjaan yang sama, dengan skor penilaian yang sudah ditetapkan oleh perusahaan yang ditunjukkan pada Tabel 1.

Tabel 1. Nilai Variabel

| Kode | Variabel | Nilai | Sub Kreteria |
|------|--------------------|-------|--------------|
| c1 | Kinerja | 100 | Sangat Baik |
| | | 75 | Baik |
| | | 50 | Cukup |
| | | 25 | Kurang |
| c2 | Pengalaman Bekerja | 100 | > 3tahun |
| | | 75 | > 2tahun |
| | | 50 | >1 Tahun |
| | | 25 | 1 Tahun |
| C3 | Loyalitas | 100 | Sangat Baik |
| | | 75 | Baik |

| | | | |
|----|--------------------|-----|-------------|
| | | 50 | Cukup |
| | | 25 | Kurang |
| c4 | Kedisiplinan | 100 | Sangat Baik |
| | | 75 | Baik |
| | | 50 | Cukup |
| | | 25 | Kurang |
| c5 | Tingkat Pendidikan | 100 | S2 |
| | | 75 | S1 |
| | | 50 | D3 |
| | | 25 | SMA/SMK |

Berikut adalah langkah-langkah penentuan indikator prioritas dalam penerimaan karyawan tetap dengan menggunakan metode *Decision Tree C4.5*. Hal yang paling terpenting dalam penyelesaian kasus pohon keputusan adalah ketersediaan data *training* atau data histori. Data yang dikumpulkan antara lain data karyawan tidak tetap yang dapat menentukan tingkat pemilihan kelas.

Tabel 2. Data Training

| No. | Nik | Nama | C1 | C2 | C3 | C4 | C5 | Keterangan |
|-------|---------|-------|-------|-------|-------|-------|-------|------------|
| 1 | 2104101 | K01 | 50 | 25 | 100 | 50 | 50 | TIDAK |
| 2 | 2105102 | K02 | 75 | 25 | 75 | 75 | 75 | YA |
| 3 | 2106103 | K03 | 25 | 100 | 75 | 100 | 25 | YA |
| 4 | 2107104 | K04 | 100 | 50 | 25 | 50 | 100 | YA |
| 5 | 2108105 | K05 | 25 | 50 | 50 | 75 | 25 | TIDAK |
| | | | | | | | | |
| | | | | | | | | |
| 150 | 2253250 | K150 | 100 | 100 | 100 | 75 | 100 | YA |

Data tabel 2 merupakan data yang digunakan dalam proses data mining merupakan data yang telah dianalisa sebelumnya dan data yang dipilih sebagai data sample, akan dilakukan transformasi data pada data yang telah dipilih sehingga sesuai dengan proses data mining dalam bentuk *MS. Excel* karena data akan digunakan dalam *tools rapid miner*.

3.3 Pengujian Data dan Hasil

Pada tahap ini dilakukan pengujian data training sesuai tujuan penelitian yaitu untuk menerapkan teknik klasifikasi menggunakan metode *decision tree* yaitu dengan konsep algoritma C4.5 dan *tools rapid miner*. Dari data training akan dibentuk suatu model pohon yang nanti akan menghasilkan sejumlah aturan dalam pohon tersebut. Model pohon akan terbentuk dengan menggunakan *tools rapid miner*. Berikut ini adalah proses pengolahan data menggunakan *tools rapid*

miner dengan beberapa sampel data testing sejumlah 15 Data karyawan baru.

Tabel 3. Data Testing

| No. | Nik | Nama | C1 | C2 | C3 | C4 | C5 |
|-------|---------|-------|-------|-------|-------|-------|-------|
| 1 | 2301001 | K01 | 100 | 25 | 75 | 100 | 100 |
| 2 | 2301002 | K02 | 25 | 25 | 100 | 25 | 25 |
| 3 | 2301003 | K03 | 75 | 100 | 75 | 100 | 100 |
| 4 | 2301004 | K04 | 25 | 50 | 100 | 50 | 100 |
| | | | | | | | |
| | | | | | | | |
| 15 | 2301015 | K15 | 25 | 75 | 100 | 25 | 75 |

Dari pengujian yang dilakukan akan mendapatkan sebuah *path* dalam pengambilan keputusan sehingga akan didapatkan nama karyawan yang diterima sebagai karyawan tetap serta dapat mengetahui indikator utama yang mempengaruhi keputusan dalam penerimaan pegawai tidak tetap menjadi karyawan tetap, dengan melakukan teknik *sorting* dari hasil analisis data testing seperti di bawah ini

. Tabel 4. Hasil Uji Data Testing

| No. | Nama | Confidence (Tidak) | Confidence (Ya) | Prediction | C1 | C2 | C3 | C4 | C5 |
|-----|------|--------------------|-----------------|------------|-----|-----|-----|-----|-----|
| 1 | K01 | 0,00 | 1,00 | YA | 100 | 25 | 75 | 100 | 100 |
| 2 | K02 | 1,00 | 0,00 | TIDAK | 25 | 25 | 100 | 25 | 25 |
| 3 | K03 | 0,00 | 1,00 | YA | 75 | 100 | 75 | 100 | 100 |
| 4 | K04 | 0,11 | 0,89 | YA | 25 | 50 | 100 | 50 | 100 |
| 5 | K05 | 0,00 | 1,00 | YA | 50 | 50 | 100 | 75 | 25 |
| 6 | K06 | 1,00 | 0,00 | TIDAK | 25 | 100 | 75 | 25 | 25 |
| 7 | K07 | 0,11 | 0,89 | YA | 50 | 75 | 75 | 50 | 75 |
| 8 | K08 | 1,00 | 0,00 | TIDAK | 50 | 25 | 75 | 25 | 50 |
| 9 | K09 | 0,00 | 1,00 | YA | 100 | 100 | 100 | 100 | 75 |
| 10 | K10 | 0,97 | 0,03 | TIDAK | 75 | 75 | 50 | 75 | 50 |
| 11 | K11 | 0,00 | 1,00 | YA | 100 | 75 | 100 | 75 | 25 |
| 12 | K12 | 0,00 | 1,00 | YA | 100 | 25 | 100 | 25 | 100 |
| 13 | K13 | 1,00 | 0,00 | TIDAK | 25 | 75 | 100 | 50 | 25 |
| 14 | K14 | 0,00 | 1,00 | YA | 50 | 100 | 50 | 25 | 100 |
| 15 | K15 | 0,11 | 0,89 | YA | 25 | 75 | 100 | 25 | 75 |

Tabel 4 merupakan hasil uji coba data testing (Data Baru) dengan *tools Rapid Miner*, dimana didapatkan data nilai pada masing-masing karyawan, yang memiliki nilai *confidence* (Tidak), nilai *confidence* (Ya), untuk nilai konvensi terjadinya *confidence* lebih umum ditulis menggunakan nilai 0 sampai 1. (Han J, 2001). dengan perbandingan ketentuan >=

maka bisa diprediksikan “Ya”, contoh: ketika nilai *confidence* (Ya) 0,5 dan *confidence* (Tidak) 0,5 maka prediksinya menunjukkan “Ya”, *Prediction* dengan keterangan “Ya” (Di terima) dan “Tidak” (Tidak diterima), untuk memudahkan dalam pemilihan karyawan yang diterima secara tepat, penulis menggunakan teknik *sorting* yang ditunjukkan pada tabel 5.

Tabel 5. Hasil *Sorting*

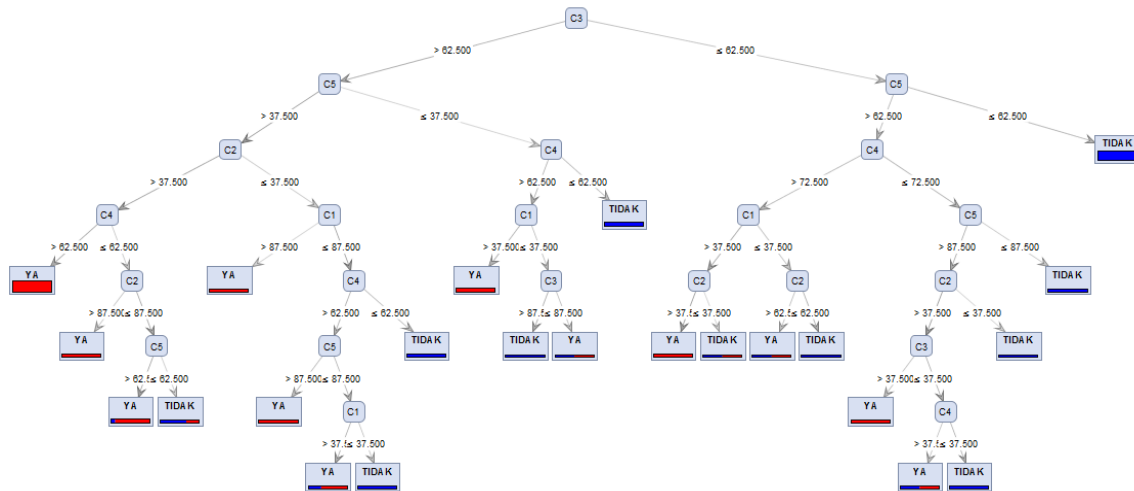
| No. | Nama | Confidence (Tidak) | Confidence (Ya) | Prediction | C1 | C2 | C3 | C4 | C5 |
|-----|------|--------------------|-----------------|------------|-----|-----|-----|-----|-----|
| 1 | K01 | 0,00 | 1,00 | YA | 100 | 25 | 75 | 100 | 100 |
| 3 | K03 | 0,00 | 1,00 | YA | 75 | 100 | 75 | 100 | 100 |
| 4 | K04 | 0,11 | 0,89 | YA | 25 | 50 | 100 | 50 | 100 |
| 5 | K05 | 0,00 | 1,00 | YA | 50 | 50 | 100 | 75 | 25 |
| 7 | K07 | 0,11 | 0,89 | YA | 50 | 75 | 75 | 50 | 75 |
| 9 | K09 | 0,00 | 1,00 | YA | 100 | 100 | 100 | 100 | 75 |
| 11 | K11 | 0,00 | 1,00 | YA | 100 | 75 | 100 | 75 | 25 |
| 12 | K12 | 0,00 | 1,00 | YA | 100 | 25 | 100 | 25 | 100 |
| 14 | K14 | 0,00 | 1,00 | YA | 50 | 100 | 50 | 25 | 100 |
| 15 | K15 | 0,11 | 0,89 | YA | 25 | 75 | 100 | 25 | 75 |
| 2 | K02 | 1,00 | 0,00 | TIDAK | 25 | 25 | 100 | 25 | 25 |
| 6 | K06 | 1,00 | 0,00 | TIDAK | 25 | 100 | 75 | 25 | 25 |
| 8 | K08 | 1,00 | 0,00 | TIDAK | 50 | 25 | 75 | 25 | 50 |
| 10 | K10 | 0,97 | 0,03 | TIDAK | 75 | 75 | 50 | 75 | 50 |
| 13 | K13 | 1,00 | 0,00 | TIDAK | 25 | 75 | 100 | 50 | 25 |

Dari tabel 5 dapat diketahui karyawan yang akan diterima sebagai pegawai tetap secara keseluruhan dengan

mengelompokkan data karyawan yang memiliki prediksi “Ya” lalu di *sorting* atau *ranking* sehingga perusahaan dapat

mengambil keputusan dalam penerimaan pegawai sesuai kebutuhan yang ada, dan

didapatkan sebuah pohon keputusan seperti gambar 3.



Gambar 3. Penentuan Indikator Utama

Keterangan Kondisi:

If $c3 \leq 62.500$ and $c5 \leq 62.500$ then tidak

If $c3 \leq 62.500$ and $c5 > 62.500$ and $c4 \leq 72.500$ and $c5 \leq 87.500$ then tidak

If $c3 \leq 62.500$ and $c5 > 62.500$ and $c4 \leq 72.500$ and $c5 > 87.500$ and $c2 \leq 37.500$ then tidak

If $c3 \leq 62.500$ and $c5 > 62.500$ and $c4 \leq 72.500$ and $c5 > 87.500$ and $c2 > 37.500$ and $c3 \leq 37.500$ and $c4 \leq 37.500$ then tidak

If $c3 \leq 62.500$ and $c5 > 62.500$ and $c4 > 72.500$ and $c1 \leq 37.500$ and $c2 \leq 62.500$ then tidak

If $c3 \leq 62.500$ and $c5 > 62.500$ and $c4 > 72.500$ and $c1 > 37.500$ and $c2 \leq 37.500$ then tidak

If $c3 > 62.500$ and $c5 \leq 37.500$ and $c4 \leq 62.500$ then tidak

If $c3 > 62.500$ and $c5 \leq 37.500$ and $c4 > 62.500$ and $c1 \leq 37.500$ and $c3 > 87.500$ then tidak

If $c3 > 62.500$ and $c5 > 37.500$ and $c2 \leq 37.500$ and $c1 \leq 87.500$ and $c4 \leq 62.500$ then tidak

If $c3 > 62.500$ and $c5 > 37.500$ and $c2 \leq 37.500$ and $c1 \leq 87.500$ and $c4 > 62.500$ and $c5 \leq 67.500$ and $c1 \leq 37.500$ then tidak

If $c3 > 62.500$ and $c5 > 37.500$ and $c2 > 37.500$ and $c4 \leq 62.500$ and $c2 \leq 87.500$ and $c5 \leq 62.500$ then tidak

Dari gambar 3 dapat diketahui bahwa yang menjadi indikator utama dalam penentuan penerimaan karyawan tetap adalah indikator C3 (Loyalitas) dan setiap percabangan dari tree C5 menyatakan kondisi yang harus dipenuhi dari indikator yang akan memiliki keterangan prediksi "Ya", dengan menggunakan indikator: C1,

C2, C3, C4 dan C5, setiap nilai dari indikator akan mempunyai percabangan menuju prediksi indikator selanjutnya, sampai pada target indikator. Setelah dilakukan pengujian menggunakan tools rapid miner, penulis melakukan pembuktian dengan cara menghitung

rumus persamaan (1) dan rumus persamaan (2).

Langkah 1 mencari *entropy* total dari data set, dengan cara mencari jumlah kasus yang sama dan total kasus sehingga menghasilkan *entropy* kasus dari pembagian tiap-tiap kasus dengan total kasus.

Tabel 6. *Entropy* Total Kasus

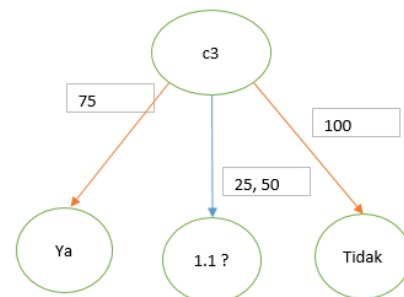
| Kasus | Jumlah K | Entropy= (-Pj*log2 Pj) |
|-------------|----------|------------------------|
| Ya | 6 | 0,970950594 |
| Tidak | 9 | |
| total Kasus | 15 | |

Langkah 2 mencari *entropy* kasus dan gain, dengan cara mencari jumlah dari nilai masing-masing atribut yang sama, dan membaginya dengan total jumlah kasus sehingga menghasilkan *entropy* nilai masing-masing atribut, untuk gain di dapat dari *entropy* total kasus dikurangi pembagian dari total jumlah kasus, total kasus, *entropy* sehingga menghasilkan gain, yang nantinya hasil dari gain akan digunakan untuk penentuan indikator.

Tabel 7. *Entropy* Masing-masing Nilai Atribut Dan Gain

| Node | Atribut | Nilai | Sum Nilai | Sum (Ya) | Sum (Tidak) | Entropi | Gain |
|------|---------|-------|-----------|----------|-------------|-------------|-------------|
| 1 | c1 | 100 | 2 | 2 | 0 | 0 | 0,260615792 |
| | | 75 | 4 | 2 | 2 | 1 | |
| | | 50 | 3 | 1 | 2 | 0,918295834 | |
| | | 25 | 6 | 1 | 5 | 0,650022422 | |
| | C2 | 100 | 4 | 3 | 1 | 0,811278124 | 0,310941626 |
| | | 75 | 2 | 0 | 2 | 0 | |
| | | 50 | 3 | 2 | 1 | 0,918295834 | |
| | | 25 | 6 | 1 | 5 | 0,650022422 | |
| | C3 | 100 | 4 | 0 | 4 | 0 | 0,570950594 |
| | | 75 | 4 | 4 | 0 | 0 | |
| | | 50 | 3 | 1 | 2 | 0,918295834 | |
| | | 25 | 4 | 1 | 3 | 0,811278124 | |
| | C4 | 100 | 2 | 2 | 0 | 0 | 0,260615792 |
| | | 75 | 4 | 2 | 2 | 1 | |
| | | 50 | 3 | 1 | 2 | 0,918295834 | |
| | | 25 | 6 | 1 | 5 | 0,650022422 | |
| | C5 | 100 | 4 | 3 | 1 | 0,811278124 | 0,253958094 |
| | | 75 | 2 | 1 | 1 | 1 | |
| | | 50 | 3 | 0 | 3 | 0 | |
| | | 25 | 6 | 2 | 4 | 0,918295834 | |

Dari pengujian gain bahwa dapat diketahui atribut yang paling berpengaruh dan tinggi yaitu C3 dengan nilai 0,57 sehingga dapat menjadi sebuah akar, Atribut C3 (loyalitas) memiliki 4 nilai, 100, 75, 50, 25, pada Atribut C3 yang ber nilai 100 sudah melakukan klasifikasi Tidak dan nilai 75 sudah melakukan klasifikasi Ya, sehingga tidak perlu dihitung dikarenakan jumlah kasus, sedangkan pada Atribut C3 yang ber nilai 50 dan 25 perlu dihitung lagi.



Gambar 4. Hasil *Entropy* dan Gain
Dapat diketahui bahwa C3 menjadi indikator utama dikarenakan memiliki nilai gain yang tertinggi.

Langkah 3 mencari *entropy* dari jumlah kasus dan gain sebagai cabang daun C3 = 25, dan menghilangkan atribut C3 untuk perhitungan gain, dikarenakan C3 sudah menjadi akar.

Tabel 8. *Entropy* Atribut C3 Dengan Nilai 25

| Kasus | Jumlah K | Entropy= $(-P_j \cdot \log_2 P_j)$ |
|-------------|----------|------------------------------------|
| Ya | 1 | 0,811278124 |
| Tidak | 3 | |
| total Kasus | 4 | |

Dari tabel 8 dapat diketahui bahwa pengujian atribut dengan jumlah kasus bernilai 25 didapatkan sebuah *entropy* kasus dengan nilai 0,811.

Tabel 9. *Entropy* Atribut C3 Dengan Nilai 25

| Node | Atribut | Nilai | Sum Nilai | Sum (Ya) | Sum (Tidak) | Entropi | Gain | | |
|------|---------|-------|-----------|----------|-------------|---------|--------------|-------------|-------------|
| 2 | c1 | 100 | 1 | 0 | 0 | 0 | 0,811278124 | 0 | |
| | | 75 | 1 | 0 | 1 | 0 | 0 | | |
| | | 50 | 1 | 1 | 0 | 0 | 0 | | |
| | | 25 | 1 | 0 | 1 | 0 | 0 | | |
| | c2 | 100 | 1 | 1 | 1 | 1 | -1 | 4,463688243 | 0,660008969 |
| | | 75 | 1 | 0 | 2 | 2 | -2 | | |
| | | 50 | 1 | 0 | 1 | 0 | 0 | | |
| | | 25 | 1 | 0 | 5 | 5 | -11,60964047 | | |
| | c4 | 100 | 0 | 2 | 0 | 0 | 0 | 2,756169493 | 0,710334802 |
| | | 75 | 0 | 2 | 2 | 2 | 0 | | |
| | | 50 | 2 | 1 | 2 | 2 | -0,29248125 | | |
| | | 25 | 2 | 1 | 5 | 5 | -3,597301488 | | |
| | c5 | 100 | 1 | 3 | 1 | 1 | -6 | 5 | 0,7169925 |
| | | 75 | 0 | 1 | 1 | 1 | 0 | | |
| | | 50 | 1 | 0 | 3 | 3 | -4,754887502 | | |
| | | 25 | 2 | 2 | 4 | 4 | -3 | | |

Dari tabel. 9 dapat diketahui bahwa *entropy* atribut C3 dengan nilai 25 menghasilkan sebuah gain, sehingga gain yang tertinggi C5 = 4,12 dijadikan sebagai cabang yang dapat dilihat pada gambar 2. =50, dan menghilangkan atribut C3 untuk perhitungan gain, dikarenakan C3 sudah menjadi akar.

Langkah 3 mencari *entropy* dari jumlah kasus dan gain sebagai cabang daun C3

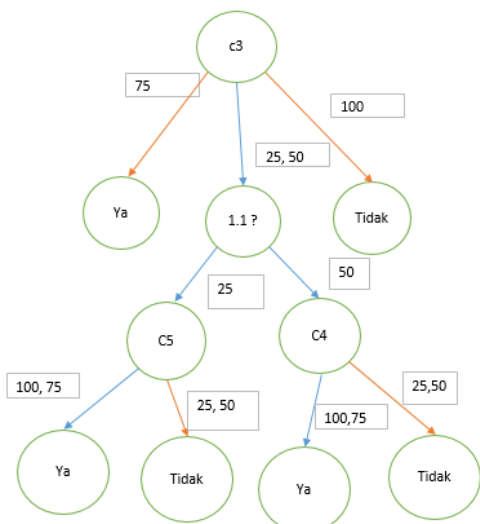
Tabel 10. *Entropy* Atribut C3 Dengan Nilai 50

| Kasus | Jumlah K | Entropy= $(-P_j \cdot \log_2 P_j)$ |
|-------------|----------|------------------------------------|
| Ya | 1 | 0,918295834 |
| Tidak | 2 | |
| total Kasus | 3 | |

Tabel 11. Entropy Atribut C3 Dengan Nilai 50

| Node | Atribut | Nilai | Sum Nilai | Sum (Ya) | Sum (Tidak) | Entropi | Gain | |
|------|---------|-------|-----------|----------|-------------|--------------|-------------|-------------|
| 3 | c1 | 100 | 0 | 0 | 0 | 0 | 0,918295834 | 0 |
| | | 75 | 1 | 0 | 1 | 0 | | |
| | | 50 | 1 | 1 | 0 | 0 | | |
| | | 25 | 1 | 0 | 1 | 0 | | |
| | C2 | 100 | 1 | 1 | 1 | 0 | 0,918295834 | 0,660008969 |
| | | 75 | 0 | 0 | 2 | 0 | | |
| | | 50 | 1 | 0 | 1 | 0 | | |
| | | 25 | 1 | 0 | 5 | 0 | | |
| | C4 | 100 | 0 | 2 | 0 | 0 | 4,121509326 | 0,710334802 |
| | | 75 | 1 | 2 | 2 | -4 | | |
| | | 50 | 0 | 1 | 2 | 0 | | |
| | | 25 | 2 | 1 | 5 | -2,804820237 | | |
| | C5 | 100 | 1 | 3 | 1 | -4,754887502 | 2,218947501 | 0,7169925 |
| | | 75 | 0 | 1 | 1 | 0 | | |
| | | 50 | 1 | 0 | 3 | -4,754887502 | | |
| | | 25 | 1 | 2 | 4 | -10 | | |

Dari tabel 9 dapat diketahui bahwa *entropy* atribut C3 dengan nilai 50 menghasilkan sebuah gain, sehingga gain yang tertinggi C4 = 4,12 dijadikan sebagai cabang yang dapat dilihat pada gambar 2.



Gambar 5. Hasil Entropy dan Gain dari Masing-masing Daun Cabang

Dapat diketahui bahwa C3 dengan nilai 50 dan 25 merupakan cabang utama yang akan menentukan cabang-cabang lainnya.

KESIMPULAN DAN SARAN

Pengambilan keputusan dalam penentuan pegawai tetap, dapat dipengaruhi oleh beberapa variabel, dalam kasus penelitian yang sudah dibahas bahwa dengan pengalaman kerja atau lamanya bekerja yang sama tidak dapat digunakan untuk tolak ukur penentuan karyawan tetap, karena didapatkan dari hasil *rule* menunjukkan bahwa Loyalitas yang

memiliki pengaruh paling besar dalam penentuan. Dari pembahasan di atas dapat disimpulkan bahwa dengan menggunakan algoritma C4.5 dan pengujian metode menggunakan *X-Fold Cross Validation* dapat membantu perusahaan untuk membuat sebuah sistem pendukung keputusan pengangkatan pegawai tidak tetap menjadi pegawai tetap, karena didapatkan indikator utama dalam penentuan karyawan tetap. Dengan adanya sistem pendukung keputusan menggunakan *Decission Tree C4.5* dapat meminimalisir kecurangan yang dilakukan oleh oknum nakal, dan perusahaan akan mendapatkan pegawai berkualitas karena pegawai berprestasi baik yang akan terpilih sebagai pegawai tetap. Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan yang diberikan, maka ada saran atau usul yang diberikan antara lain:

1. Untuk peneliti selanjutnya dapat dibandingkan dengan metode *bayesian*, dengan pembobotan variabel untuk mendapatkan *accuracy* yang lebih.
2. Untuk peneliti selanjutnya agar dapat mengembangkan penelitian ini dengan mengoptimalkan hasil dari algoritma C4.5 menggunakan metode *bootstrap* guna mempersingkat waktu dalam komputasi, sehingga waktu yang akan digunakan dalam komputasi penentuan karyawan menjadi lebih efisien.

DAFTAR PUSTAKA

- Cahyo Dimas K, H. D. (2017). Sistem Pendukung Keputusan Menggunakan ID3. *AITI*, 13-30.
- H. Jantan, A. R. (2011). Towards applying Data Mining Techniques for Talent Mangement . *International Conference on Computer Engineering and Applications*, 2, 476-481.
- Hamidah Jantan, A. R. (2010). Human Talent Prediction in HRM using C4.5. *International Journal on Computer Science and Engineering*, 2, 2526-2534 .
- Han J, K. M. (2001). Data Mining Concepts and Techniques. *Morgan Kaufmann*, California.
- John M. Kirimi, C. A. (2016). Application of Data Mining Classification in Employee. *International Journal of Computer Applications* , 146, 28-35.
- Julianto, W., Yinitarini, R., & Sophan, M. K. (2014). Algoritma C4.5 Untun Penilaian Kinerja Karyawan. *SCAN*, 8(ISSN : 1978-0087), 33-39.
- Mustofa, Z., & Suasana, I. S. (2018). Algoritma Clustering K-Medoids Pada E-Government Bidang Information And Communication Technology Dalam Penentuan Status Edgi. *Jurnal Teknologi Informasi Dan Komunikasi*, Vol 9 No 1(ISSN: 2598-9707), 1-10.
- N. Sirvaram, K. R. (2010). Applicability of Clustering and Classification Algorithms. *International Journal of Computer Applications*, 4, 24-28.
- Pandiangan, P. M. (2017). Penentuan Indikator Prioritas Pembangunan Desa Menggunakan Metode seleksi Fitur. *Institut Teknologi Sepuluh Nopember*. Surabaya.
- Qasem A. Al-Radaideh, E. A. (2012). Using Data Mining Techniques to Build a Classification Model for Predicting Employees . *International Journal of Advanced Computer Science and Applications*, 3, 144-151.
- Sembiring, I. (2004). Piranti Bantu Pendukung Keputusan Kelayakan Investasi E-Government. *Universitas Gadjah Mada*.