

Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE

Fandy Ferdian Harryanto*¹, Seng Hansun²

^{1,2}Universitas Multimedia Nusantara; Jl. Scientia Boulevard, Gading Serpong, Tangerang, Banten-15811 Indonesia
e-mail: *¹vhanz54@yahoo.com, ²hansun@umn.ac.id

Abstrak

Perusahaan pada umumnya memerlukan pegawai yang memiliki kemampuan yang baik, perilaku yang baik serta dapat menyelesaikan pekerjaan yang diberikan kepadanya. Namun terdapat beberapa kesulitan untuk mengetahui kualitas orang-orang yang memiliki potensi baik sebagai pegawai pada suatu perusahaan. Oleh karena itu diperlukan cara atau metode untuk mengidentifikasi calon pegawai suatu perusahaan. Algoritma C4.5 dapat digunakan untuk melakukan prediksi dan klasifikasi terhadap calon pegawai yang berpotensi untuk masuk ke dalam perusahaan dengan cara membuat pohon keputusan berdasarkan data-data yang sudah ada dan melakukan prediksi terhadap calon pegawai baru yang ingin masuk ke perusahaan. Berdasarkan metode pengukuran akurasi ten-fold cross validation telah didapatkan hasil pengukuran tingkat keberhasilan prediksi calon pegawai baru sebesar 71% dengan menggunakan aplikasi prediksi calon pegawai yang menerapkan algoritma C4.5.

Kata kunci: Algoritma C4.5, Calon Pegawai, Prediksi, Pohon Keputusan.

Abstract

A company in general needs employee that have good ability, good manners and also can company. But there are difficulties in finding the qualities of people as a good employee candicacy. That's why we need a way or method to identify peoples with the potential to become a new employee candidate. C4.5 algorithm can be used to predict and classify new employee candidate that have the potential to get into a corporation by using decision tree according to the data that we have and predict the new employee candidate qualities. According to the testing method called ten-fold cross validation, the accuracy of the prediction for the new employee candidate is 71% by using the built prediction application which implementing C4.5 algorithm.

Keywords: C4.5 Algorithm, Decision Tree, Employee Candidate, Prediction.

1. PENDAHULUAN

Pegawai merupakan sumber daya yang paling penting dalam suatu perusahaan dan pegawai yang baik dan memenuhi standar perusahaan hanya dapat diperoleh melalui proses penerimaan pegawai yang efektif [1]. Pegawai yang baik dan berkualitas tentu saja memiliki dampak yang positif terhadap perusahaan tempat mereka bekerja. Oleh karena itu perusahaan harus menyeleksi orang-orang yang ingin masuk dan bekerja ke dalam perusahaan itu dengan baik.

Penerimaan calon pegawai baru merupakan sebuah tahap dimana sebuah perusahaan

melakukan rekrutmen terhadap orang-orang yang melamar ke perusahaan tersebut dan menentukan apakah orang tersebut memenuhi kriteria dan kebutuhan unit kerja pada perusahaan tersebut. Menurut Tjahyono dan Anggara [2] alasan dilakukannya penerimaan calon pegawai baru adalah pengembangan unit usaha perusahaan tersebut yang mengakibatkan perlunya penambahan pegawai baru dan kebutuhan perusahaan tersebut untuk mengisi posisi kosong yang ditinggalkan oleh pegawai lama mereka.

Selain itu, Slamet [3] menyatakan bahwa proses rekrutmen ini berguna untuk mendapatkan informasi mengenai kecakapan, kepribadian, dan kemampuan lain yang dimiliki oleh tenaga kerja yang melamar ke perusahaan tersebut. Informasi tersebut dipandang sangat diperlukan untuk menentukan apakah pekerja tersebut berkualitas dan pantas untuk masuk ke dalam perusahaan. Tenaga kerja yang berkualitas tentunya akan membantu untuk meningkatkan kinerja perusahaan.

Berdasarkan hasil wawancara dengan salah satu pekerja pada PT WISE, Bapak Paulus Harryanto, dapat ditarik kesimpulan bahwa perusahaan ini masih membutuhkan bantuan tambahan pada proses perekrutan calon pegawai. Menurut Bapak Paulus masih terdapat beberapa pegawai kantor yang malas dan tidak kompeten. Hal ini berarti proses perekrutan pegawai pada perusahaan ini tidak berjalan dengan baik. Selain itu, berdasarkan hasil wawancara dengan Cendrawati selaku divisi *Human Resource Development* (HRD), ada beberapa faktor yang menentukan apakah seseorang dapat diterima ke dalam perusahaan. Faktor-faktor tersebut adalah: umur, pendidikan terakhir, pengalaman bekerja, jenis kelamin, perilaku saat wawancara, permintaan gaji awal dan penyakit yang sedang diderita.

Algoritma C4.5 dapat digunakan untuk meneliti berbagai macam hal, diantaranya adalah prediksi tingkat kemenangan pada pertandingan bola [4], mencari pola prediksi hujan [5], hingga menentukan guru terbaik [6]. Sebelumnya juga telah ada yang melakukan penelitian yang serupa dengan menggunakan algoritma ini, namun atribut yang digunakan untuk mengklasifikasikan pohon keputusan berbeda. Adapun pembedanya lainnya adalah penelitian tersebut dilakukan kepada calon pegawai negeri sipil. Penelitian tersebut dilakukan oleh Kumara dan Supriyanto [7] dengan judul “Klasifikasi Data Mining untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5”. Tingkat akurasi yang didapatkan menggunakan algoritma C4.5 ini sudah cukup tinggi, oleh karena itu dapat disimpulkan algoritma ini cocok untuk diimplementasikan pada penelitian yang melibatkan proses perekrutan [7].

Algoritma C4.5 digunakan pada penelitian ini untuk memprediksi proses penerimaan calon pegawai baru pada PT WISE. Menurut penelitian HSSINA, dkk.[8], algoritma C4.5 merupakan algoritma terkuat untuk proses pembuatan *decision tree* jika dibandingkan dengan algoritma lain seperti ID3, C5.0, dan CART. Berdasarkan penelitian inilah algoritma C4.5 diputuskan untuk digunakan pada penelitian ini karena algoritma ini lebih kuat daripada algoritma lainnya. Data pegawai yang bekerja sebagai pegawai lapangan (*maintenance, service*) akan digunakan sebagai *data training* untuk pembuatan *decision tree* yang akan digunakan untuk memprediksi penerimaan pegawai dan mengetes tingkat keberhasilan prediksi tersebut menggunakan *data testing* terhadap pegawai yang memiliki jabatan pekerjaan yang sama.

2. LANDASAN TEORI

Beberapa literatur yang mendasari penelitian ini akan dibahas pada bab ini, meliputi sistem pendukung keputusan, pohon keputusan, algoritma C4.5, dan validasi silang.

2.1 Sistem Pendukung Keputusan

Sistem pendukung keputusan adalah sebuah sistem yang dapat membantu manusia untuk mengambil keputusan secara objektif. Konsep sistem seperti ini pertama kali dicetuskan pada tahun 1970-an oleh Michael S.Scott Morton, Michael pertama kali menyebut sistem seperti

ini dengan nama *Management Decision System*. Maksud dan tujuan dari adanya sistem pengambil keputusan adalah mendukung pengambil keputusan untuk memilih alternatif keputusan menggunakan model-model pengambilan keputusan dan untuk menyelesaikan masalah yang bersifat terstruktur, semi terstruktur, maupun tidak terstruktur [9].

Menurut Mengkepe [9], kerangka dasar pengambilan keputusan manajerial dalam tipe keputusan dibagi menjadi beberapa bagian, yaitu:

1. Terstruktur: Berisi masalah yang sering terjadi, solusinya dapat berupa standar dan baku.
2. Tidak Terstruktur: Berisi masalah kompleks menggunakan pemecahan masalah yang tidak standar, solusinya melibatkan intuisi manusia sebagai dasar pengambil keputusan.
3. Semi Terstruktur: Gabungan antara keputusan terstruktur dengan tidak terstruktur, solusinya merupakan gabungan antara prosedur solusi standar dengan kemampuan individu manusia.

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 [10].

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 [10].

Arsitektur pohon keputusan dibuat sedemikian rupa agar menyerupai pohon asli, dimana terdapat beberapa bagian yaitu [11]:

- *Root Node*: *Node* ini terletak pada bagian paling atas dari pohon keputusan.
- *Internal Node*: *Node* ini merupakan percabangan dimana membutuhkan satu *input* dan mengeluarkan maksimal dua *output*.
- *Leaf Node*: *Node* ini merupakan *node* yang terletak pada ujung pohon. *Node* ini hanya memiliki satu *input* dan tidak memiliki *output*.

Menurut Triisant [11], pohon keputusan ini memiliki beberapa kekurangan dan kelebihan, yakni:

1. Kelebihan Pohon Keputusan:
 - Daerah pengambilan keputusan yang kompleks dapat diubah menjadi sederhana.
 - Dapat menghilangkan perhitungan yang tidak penting karena proses pengujian hanya berdasarkan kriteria yang diperlukan saja.
 - Proses pemilihan fitur dari *internal node* yang berbeda lebih fleksibel. Fitur yang telah dipilih ini akan menjadi pembeda antara kriteria yang satu dengan kriteria lainnya.
 - Metode ini dapat menghindari munculnya permasalahan dengan cara menggunakan kriteria dengan jumlah yang sedikit pada *node internal* tanpa mengurangi kualitas keputusan yang dihasilkan.
2. Kekurangan Pohon Keputusan :
 - Dapat terjadi *overlap* apabila hasil keputusan dan kriteria yang digunakan jumlahnya sangat banyak. Hal ini juga dapat berakibat bertambahnya waktu yang digunakan untuk pengambilan keputusan dan jumlah memori yang dibutuhkan semakin tinggi.
 - Akumulasi jumlah *error* dari setiap tingkat pohon keputusan besar.
 - Mendesain pohon keputusan yang optimal sulit.
 - Kualitas keputusan yang didapatkan sangat tergantung dengan bagaimana pohon tersebut didesain.

2.3 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk melakukan proses klasifikasi data dengan menggunakan teknik pohon keputusan. Algoritma C4.5 merupakan

ekstensi dari algoritma ID3 dan menggunakan prinsip *decision tree* yang mirip [11]. Algoritma ini sudah sangat terkenal dan disukai karena memiliki banyak kelebihan. Kelebihan ini misalnya dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain [11].

Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi [12]. Kemudian secara rekursif cabang-cabang pohon diperluas sehingga seluruh pohon terbentuk. Menurut kamus IGI Global (*International Publisher of Progressive Academic*), *entropy* adalah jumlah data yang tidak relevan terhadap informasi dari suatu kumpulan data [13]. *Gain* adalah informasi yang didapatkan dari perubahan *entropy* pada suatu kumpulan data, baik melalui observasi atau bisa juga disimpulkan dengan cara melakukan partisipasi terhadap suatu *set* data [12].

Berdasarkan apa yang ditulis oleh Jefri [10], terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

1. Memilih atribut sebagai akar
3. Membuat cabang untuk masing-masing nilai
4. Membagi setiap kasus dalam cabang
5. Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama.

Menurut Jiandi [12] data yang dimiliki harus disusun menjadi sebuah tabel berdasarkan kasus dan jumlah responden sebelum dilakukan perhitungan untuk mencari nilai *entropy* dan *gain*.

$$Entropy(S) = \sum_{i=0}^n -p_i * \log_2 p_i \quad (1)$$

Rumus (1) merupakan rumus yang digunakan dalam perhitungan *entropy* yang digunakan untuk menentukan seberapa informatif atribut tersebut. Berikut keterangannya [10]:

S : Himpunan kasus

n : Jumlah partisi S

p_i : Jumlah kasus pada partisi ke- i

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

Rumus (2) merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan *entropy*. Berikut keterangannya [10]:

S : Himpunan kasus

n : Jumlah partisi atribut A

$|S_i|$: Jumlah kasus pada partisi ke- i

$|S|$: Jumlah kasus dalam S

Dengan mengetahui rumus-rumus diatas, data yang telah diperoleh dapat dimasukkan dan diproses dengan algoritma C4.5 untuk proses pembuatan *decision tree*.

```

FormTree(T)
(1) ComputeClassFrequency(T);
(2) if OneClass or FewCases
    Return a leaf;
    Create a decision node N;
(3) ForEach Attribute A
    ComputeGain(A);
(4) N.test = AttributeWithBestGain;
(5) if N.test is Continuous
    Find Threshold;
(6) ForEach T1 in the Splitting of T
(7)   if T1 is Empty
        Child of N is a leaf
    Else
        Child of N = FormTree(T1);
(8) ComputeErrors of N;
    Return N;

```

Gambar 1. *Pseudocode* Algoritma C4.5 [14]

Gambar 1 merupakan *pseudocode* dari algoritma C4.5 yang berfungsi untuk pembentukan pohon keputusan. Perhitungan dimulai dari menghitung banyaknya jumlah atribut dan menentukan atribut mana yang akan digunakan sebagai akar dari pohon keputusan. Selanjutnya akan dilakukan perhitungan *entropy* dan *gain* untuk menentukan *leaf* dari pohon keputusan tersebut. Setelah semua perhitungan selesai dilakukan, pohon keputusan dapat dibentuk berdasarkan nilai *gain* yang telah dihitung sebelumnya. Atribut dengan nilai *gain* tertinggi akan terletak pada prioritas yang lebih tinggi dan memiliki kedudukan yang lebih tinggi juga pada pohon keputusan.

2.4 Validasi Silang

Validasi silang merupakan metode statistika yang digunakan untuk melakukan evaluasi dan komparasi terhadap sebuah satu *set* data dengan cara membagikan data tersebut menjadi dua bagian, yaitu *data training* dan *data testing*. Salah satu jenis validasi silang adalah *ten-fold cross validation* [15]. Validasi ini dilakukan dengan cara membagi suatu set data menjadi sepuluh segmen $d1 - d10$ yang berukuran sama besar dengan cara melakukan pengacakan data. Kemudian $d1$ akan digunakan terlebih dahulu untuk proses *training* dan dilakukan validasi menggunakan sisa dari data selain $d1$. Setelah itu $d2$ akan digunakan untuk *training*, sementara sisa data selain $d2$ digunakan untuk validasi, dan seterusnya. Dengan melakukan validasi seperti ini maka akurasi yang akan didapatkan akan lebih tinggi [15].

3. METODE PENELITIAN

Implementasi algoritma C4.5 untuk prediksi penerimaan calon pegawai baru menggunakan langkah-langkah penelitian sebagai berikut:

1. Studi Pustaka

Penelitian dimulai dengan mempelajari informasi serta algoritma yang bersangkutan dengan penelitian ini dengan cara membaca *e-book*, *e-journal*, serta beberapa referensi pembelajaran lain. Pada tahapan ini konsep-konsep yang dibutuhkan dalam penelitian akan dimatangkan, seperti definisi dari algoritma C4.5 dan penerapannya.

2. Pengumpulan Sampel Data

Pada tahap ini proses pengumpulan sampel data akan dilakukan dengan cara meminta data pegawai langsung kepada perusahaan, yakni PT WISE. Data pegawai ini mencakup berbagai macam informasi mulai dari nama, umur, gaji, alamat, dan atribut lain yang kemudian akan diolah menggunakan algoritma C4.5. Jumlah data yang akan digunakan adalah 84 data pegawai lapangan.

3. Analisis Sampel Data

Setelah data-data pegawai didapatkan, dilakukan pemilahan dan perhitungan atribut data sesuai dengan parameter yang telah ditentukan untuk dilakukan perhitungan nilai *entropy* dan *gain* untuk mendapatkan gambaran umum dari suatu *set* data.

4. Perancangan dan Pembangunan Aplikasi

Pada tahap ini, beberapa hal akan ditentukan, yakni prosedur dan proses apa saja yang dapat dilakukan oleh aplikasi, alur proses, serta tampilan dasar aplikasi. Rancangan aplikasi akan direpresentasikan dalam bentuk diagram yang akan menggambarkan alur proses dari aplikasi. Setelah itu barulah aplikasi mulai dibangun menggunakan bahasa pemrograman yang tepat.

5. Uji Coba Aplikasi

Pada tahap ini uji coba akan dilakukan terhadap aplikasi yang sudah dibuat terhadap data-data yang telah dikumpulkan sebelumnya. Pengamatan apakah algoritma C4.5 dapat terimplementasi dengan baik pada sistem serta dapat memiliki tingkat akurasi yang tinggi pada proses prediksi calon pegawai baru juga dilakukan pada tahap ini.

6. Analisis Hasil Aplikasi

Setelah uji coba pada aplikasi berhasil dilakukan, selanjutnya akan dilakukan pengukuran tingkat akurasi prediksi menggunakan validasi silang *ten-fold cross validation*. Validasi silang digunakan agar akurasi pengukuran prediksi lebih tepat.

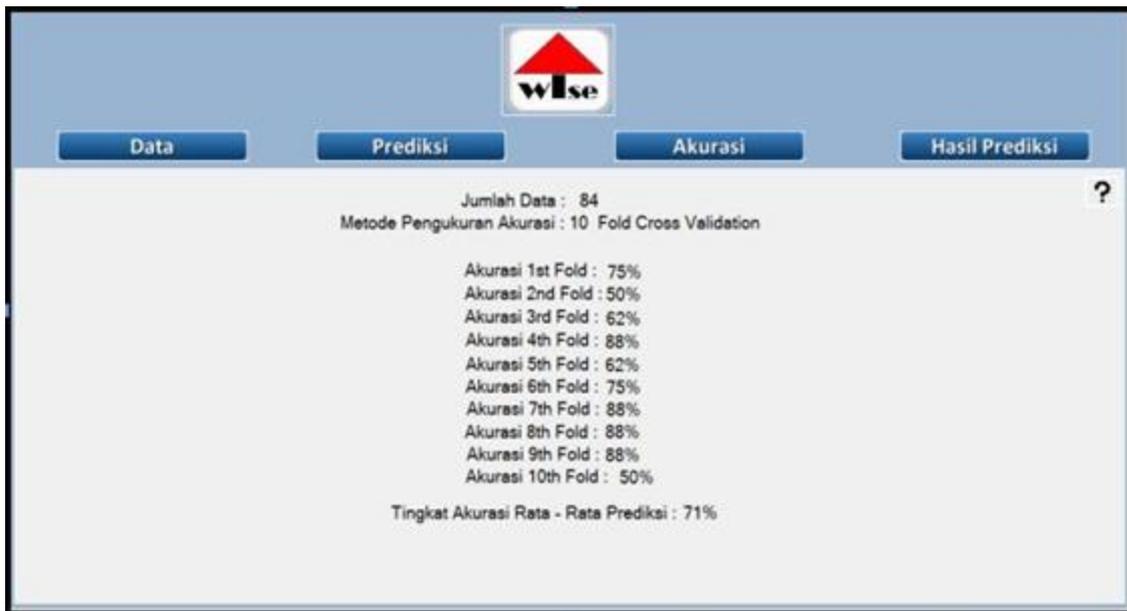
7. Penulisan Laporan

Pada tahap ini akan dilakukan proses pencatatan setiap kegiatan yang dilakukan selama proses penelitian ini berlangsung dan menyalurkan informasi tersebut ke dalam bentuk laporan sebagai bentuk dokumentasi.

4. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil dari pengukuran akurasi aplikasi yang telah dibangun dalam penelitian. Uji coba dilakukan menggunakan data *sample* sebanyak 84 dan menggunakan metode pengukuran *ten fold cross validation*. Data tersebut kemudian akan dibagi menjadi 10 buah kelompok *data training* dan *data testing*. Setelah itu barulah uji coba untuk mengukur akurasi aplikasi dilakukan.

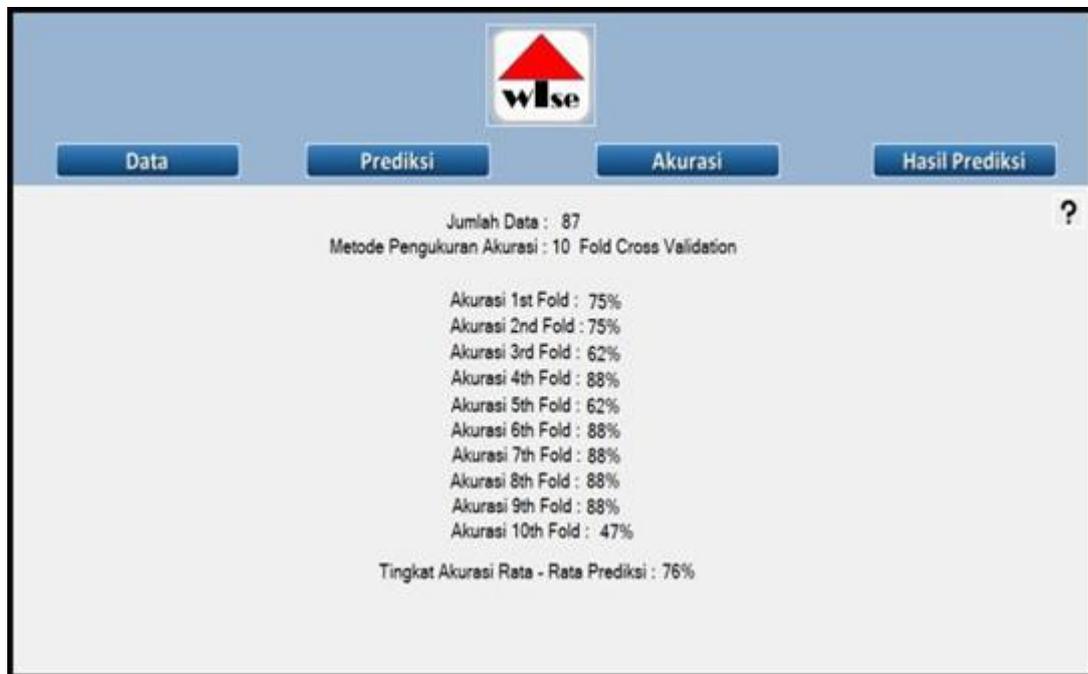
Analisis yang akan dilakukan pada aplikasi adalah perhitungan tingkat akurasi menggunakan metode *10-fold cross validation*, baik akurasi perkluster maupun secara keseluruhan. Uji coba akan dilakukan pada 84 buah sampel data yang ada pada *database* secara otomatis melalui sistem.



Gambar 2. Hasil Pengukuran Akurasi

Gambar 2 menjabarkan hasil perhitungan dari *cross validation* yang dilakukan oleh aplikasi. Dapat dilihat hasil perhitungan akurasi untuk tiap kluster dan secara keseluruhan.

Hasil pengukuran akurasi secara keseluruhan adalah 71%. Hasil prediksi tidak mencapai angka 100% karena ada beberapa kasus dimana sampel data masih kurang banyak sehingga hasil prediksi masih bersifat ambigu.



Gambar 3. Hasil Pengukuran Akurasi dengan Tambahan

Dapat dilihat pada Gambar 3 ketika perhitungan dilakukan dengan penambahan tiga sampel data baru, maka perhitungan akurasi meningkat sebanyak 5%. Hal ini berarti beberapa

sampel tambahan tersebut dapat memperbaharui pohon sehingga mengurangi tingkat kesalahan prediksi.

5. KESIMPULAN

Implementasi algoritma C4.5 untuk melakukan prediksi terhadap calon pegawai baru pada PT WISE telah berhasil dilakukan. Hasil tingkat keberhasilan prediksi calon pegawai baru di PT WISE secara keseluruhan yang telah diukur menggunakan metode *ten-fold cross validation* adalah sebesar 71%. Untuk penelitian selanjutnya, sebaiknya dilakukan uji coba menggunakan sampel data yang lebih banyak agar tingkat akurasi aplikasi dapat ditingkatkan.

DAFTAR PUSTAKA

- [1] IGI Global Dictionary, 2015, *What is Information Gain*, <http://www.igiglobal.com/dictionary/information-gain/14407/>, diakses tgl 10 Desember 2015.
 - [2] Jiandi, R., 2016, *Implementasi Algoritma C4.5 untuk Prediksi Potensi Mahasiswa Sebagai Pengurus Organisasi Menggunakan Data Hasil PAPI KOSTICK (Studi Kasus: Universitas Multimedia Nusantara)*, Universitas Multimedia Nusantara, Tangerang.
 - [3] Jefri, 2013, *Implementasi Algoritma C4.5 Dalam Aplikasi untuk Memprediksi Jumlah Mahasiswa yang Mengulang Mata Kuliah STMIK Yogyakarta*, STMIK AMIKOM, Yogyakarta.
 - [4] Kumara, R. dan Supriyanto, C., 2015, *Klasifikasi Data Mining untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5*, Universitas Dian Nuswantoro, Jawa Tengah.
 - [5] Mengkepe, E., 2004, *Sistem Pendukung Keputusan Pemberian Kredit Mobil PT. Astra International Tbk., Isuzu Division Makassar*, Universitas Widyatama, Bandung.
 - [6] Putri, S. U., 2015, *Implementasi Metode C4.5 untuk Menentukan Guru Terbaik pada SMK 1 Percut Sei Tuan Medan*, STMIK Budi Darma, Medan.
 - [7] HSSINA, B., dkk., 2014, *A Comparative Study of Decision Tree ID3 and C4.5*, Sultan Moulay Slimane University, Morocco.
 - [8] Raditya, A., 2012, *Implementasi Data Mining Classification untuk Mencari Pola Prediksi Hujan dengan Menggunakan Algoritma C4.5*, Universitas Gunadarma, Depok.
 - [9] Sulistiyani, T. dan Ambar, R., 2003, *Manajemen Sumber Daya Manusia*, Graha Ilmu, Yogyakarta.
 - [10] Slamet, A., 2007, *Manajemen Sumber Daya Manusia*, Universitas Negeri Semarang, Semarang.
 - [11] Tjahyono, A. dan Anggara, A. M., 2010, *Sistem Pendukung Keputusan Penerimaan Pegawai Baru pada PT. Kanasritex Semarang, Techno.com, Vol. 9 No.3*.
-

-
- [12] Marwana, 2014, *Algoritma C4.5 untuk Simulasi Prediksi Kemenangan Dalam Pertandingan Sepakbola*, STIMED, Nusa Palapa, Makassar.
 - [13] Triisant, 2015, *Pohon Keputusan dengan Algoritma C4.5*, <http://dokumen.tips/documents/algoritma-c45.html>, Diakses tgl 21 Maret 2016.
 - [14] Ruggieri, S., 2002, Efficient C4.5, *IEEE Transaction on Knowledge and Data Engineering* 14(2), hal.438-444.
 - [15] Refaeilzadeh, P., Tang, L., dan Liu, H., 2009, *Cross-Validation*, *Encyclopedia of Database Systems*, hal.532-538.
-