

Prediksi Penerimaan Calon Karyawan Dengan Menggunakan Algoritma C4.5 Pada Biro Kesejahteraan Rakyat Provinsi Sumatera Utara

Amir Mahmud Husein^{1*}, Mahendra Brutu²

^{1,2}Universitas Prima Indonesia, Medan

¹amirmahmud@unprimdn.ac.id



Histori Artikel:

Diajukan: 25 Sep 2022

Disetujui: 29 Sep 2022

Dipublikasi: 29 Sep 2022

Kata Kunci:

Algoritma C4.5, Data Mining, Decision Tree, Prediksi, Prediksi Calon Karyawan,

Digital Transformation

Technology (Digitech) is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak

Untuk menentukan calon karyawan yang memenuhi kriteria maka dibutuhkan rekomendasi yang tepat. Perekomendasi yang tepat membutuhkan jumlah data yang banyak, Algoritma C4.5 merupakan salah satu metode yang dapat menganalisis data dalam jumlah yang banyak atau biasa disebut data mining. Algoritma C4.5 digunakan pada penelitian ini untuk memprediksi proses penerimaan calon karyawan baru dengan cara mengklasifikasikan data calon karyawan berdasarkan kriteria yang ditentukan. Pada penelitian ini dibuat suatu aplikasi yang mengimplementasikan pohon keputusan dengan algoritma C4.5, data yang digunakan dalam penelitian ini adalah data calon pegawai. Algoritma C 4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan. Adapun alur proses pengerjaan penelitian ini dimulai dari Pengumpulan Data sesuai dengan beberapa metode yang telah dijelaskan sebelumnya. Kemudian setelah dataset terkumpul, dilakukan tahap *preprocessing* untuk proses pengolahan data. Kemudian dilakukan proses perhitungan baik manual maupun menggunakan sistem dari RapidMiner sebagai perbandingan hasil ke akuratan. Hingga pada akhirnya diperoleh hasil akhir. Setelah pengujian dilakukan beberapa kali maka operator yang ditetapkan adalah dengan menggunakan operator *split validation* dan menggunakan *Stratified Sampling* dengan training data 90% dan testing data 10% dengan tingkat akurasi sebesar 81,82% dan classification error sebesar 18,18%.

PENDAHULUAN

Semakin meningkatnya tingkat pelamar kerja menyebabkan sulitnya perusahaan atau instansi dalam memilih karyawan yang kompeten sesuai dengan kebutuhan pada bidangnya masing-masing. Dalam persaingan perusahaan di masa kini, maka perusahaan memerlukan sumber daya manusia yang baik dan dapat diandalkan. Untuk itu, maka diperlukan proses seleksi yang dilaksanakan dengan baik, benar, terstruktur dan akurat agar menghasilkan SDM yang bermutu dan mendukung kesuksesan perusahaan. SDM merupakan sebuah spekulasi bagi organisasi dan perusahaan untuk dapat berkembang dan berkreasi dalam memahami visi dan misi perusahaan (Sukma dkk, 2019).

Berdasarkan hasil wawancara yang dilakukan dengan salah satu karyawan di Biro Kesejahteraan Rakyat, maka didapatkan informasi mengenai permasalahan yang sedang dialami yaitu perusahaan ini masih membutuhkan bantuan tambahan dalam proses perekrutan calon karyawan. Dimana proses perekrutan calon karyawan masih belum berjalan dengan baik. Hal ini ditandai dengan adanya karyawan yang tidak berkompeten dan kurang ahli dalam bidangnya masing-masing. Selain itu, berdasarkan informasi yang didapatkan, ada beberapa kriteria yang menentukan apakah calon karyawan tersebut dapat diterima bekerja di perusahaan. Kriteria tersebut adalah: umur, pendidikan terakhir, pengalaman bekerja, jenis kelamin, perilaku saat wawancara, permintaan gaji awal dan penyakit yang sedang diderita.

Untuk menentukan calon karyawan yang memenuhi kriteria maka dibutuhkan rekomendasi yang tepat. Perekomendasi yang tepat membutuhkan jumlah data yang banyak, Algoritma C4.5 merupakan salah satu metode yang dapat menganalisis data dalam jumlah yang banyak atau biasa disebut data mining. Algoritma C4.5 digunakan pada penelitian ini untuk memprediksi proses penerimaan calon karyawan baru dengan cara mengklasifikasikan data calon karyawan berdasarkan kriteria yang ditentukan. Pada penelitian ini dibuat suatu aplikasi yang mengimplementasikan pohon keputusan dengan algoritma C4.5, data yang digunakan dalam penelitian ini adalah data calon pegawai. Algoritma C 4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan.

STUDI LITERATUR

Data mining muncul sekitar tahun 90an. *Data mining* merupakan salah satu cabang ilmu komputer yang relatif baru. *Data mining* juga merupakan istilah yang digunakan untuk menemukan pengetahuan yang

tersembunyi di dalam sebuah *database* (Hasibuan, D.Z, 2019), data mining digunakan untuk membuat sebuah prediksi yang valid dengan menggunakan berbagai alat analisis data untuk menemukan pola dan hubungan di dalam sebuah data, *Data mining* juga disebut sebagai kegiatan menemukan pola yang menarik dari data yang berjumlah besar serta penting dalam memberikan keuntungan kepada pemilik data. Data dapat disimpan dalam *database* atau penyimpanan informasi lainnya (Rahmaningsih, 2018).

Data mining bertujuan untuk menggali, menambahkan pengetahuan dari data yang sudah dimiliki. *Data mining* sering disebut *Knowledge Discovery in Database* (KDD) merupakan sebuah bidang ilmu yang banyak membahas tentang pola sebuah data. Serangkaian proses digunakan untuk mendapatkan pengetahuan dari kumpulan data disebut *data mining*. Sebuah data yang besar sering tidak berguna dan hanya akan dijadikan sampah jika tidak dimanfaatkan dengan benar. *Data mining* menyelesaikan masalah ini dengan menganalisa data yang besar kemudian membuat aturan, pola, ataupun model untuk mengenali data baru yang tidak berada pada data yang tersimpan (Maricar & Pramana, 2019). Metode dalam pengolahan data mining, yaitu (Ardiansyah dkk, 2018):

1. *Predictive modelling* yaitu pengolahan data dengan melakukan prediksi. Yang bertujuan untuk membangun model prediksi pada suatu nilai dengan ciri-ciri tertentu.
2. Asosiasi yaitu teknik yang mempelajari hubungan antar data.
3. *Clustering* yaitu pengelompokan teknik untuk mengidentifikasi data ke dalam suatu kelompok.
4. Klasifikasi yaitu teknik klasifikasi data. Klasifikasi memiliki perbedaan dengan *clustering* yaitu pada datanya. Pada *clustering* tidak ada variabel dependen, sedangkan klasifikasi harus ada variabel dependen.
5. Estimasi hampir sama dengan klasifikasi, hanya saja variabel estimasi yaitu numerik. Model dibangun dengan menggunakan data yang lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.

METODE

Dalam pengerjaan penelitian ini dilakukan beberapa metode pengumpulan data, yaitu:

1. Studi Kepustakaan
Mengumpulkan dan mempelajari jurnal penelitian nasional dan internasional yang berkaitan dengan objek penelitian.
2. Studi Lapangan
 - a. Observasi
Penulis mengamati langsung terhadap kegiatan-kegiatan yang dilakukan di Biro Kesejahteraan Rakyat agar mengetahui setiap proses yang dikerjakan pada bagian penerimaan calon karyawan.
 - b. Wawancara
Melakukan tanya jawab (*interview*) dengan pihak-pihak Biro Kesejahteraan Rakyat yang berhubungan langsung dengan masalah yang diteliti pada perusahaan.
 - c. Sampling
Pengumpulan data yang dilakukan dengan cara melakukan pengambilan data catatan yang berkaitan dengan objek penelitian 3 tahun.

Adapun alur proses pengerjaan penelitian ini dimulai dari Pengumpulan Data sesuai dengan beberapa metode yang telah dijelaskan sebelumnya. Kemudian setelah dataset terkumpul, dilakukan tahap *preprocessing* untuk proses pengolahan data. Kemudian dilakukan proses perhitungan baik manual maupun menggunakan sistem dari RapidMiner sebagai perbandingan hasil ke akuratan. Hingga pada akhirnya diperoleh hasil akhir.

HASIL

Pengujian dilakukan dengan menggunakan 2 operator yaitu operator *split validation* dan *cross validation* untuk menemukan hasil akurasi yang baik. Untuk pengujian pertama menggunakan operator *split validation*. Parameter yang digunakan pada algoritma C4.5 yaitu *maximal depth* 10, *confidence* 0,25, *minimal gain* 0,04 dan *criterion* yang digunakan *gain_ratio*. *Criterion* digunakan untuk menetapkan atribut akar dari *decision tree*, *maximal depth* digunakan untuk membatasi ukuran pohon keputusan, *confidence* digunakan untuk menentukan tingkat kepercayaan agar mengurangi kesalahan perhitungan pada pemangkasan, *minimal gain* digunakan untuk menentukan nilai *gain* minimal untuk menghasilkan simpul pohon keputusan. Untuk operator *performance* yang digunakan yaitu *accuracy*, *classification error*.

Type sampling data yang akan digunakan seperti *linear sampling*, *stratified sampling*, *shuffled sampling*, dan *automatic sampling* Pengujian dilakukan agar dapat memilih tingkat akurasi yang paling bagus, berikut tabel pengujiannya:

1. Linear Sampling

Tabel 1. Hasil *Linear Sampling*

Training Data-Testing Data	Accuracy	Classification Error
60%-40%	71,74%	28,26%
70%-30%	74,29%	25,71%
80%-20%	73,91%	26,09%
90%-10%	75,00%	25,00%

2. *Shuffled Sampling*

Tabel 2. Hasil *Shuffled Sampling*

Training Data-Testing Data	Accuracy	Classification Error
60%-40%	65,22%	34,78%
70%-30%	77,14%	22,86%
80%-20%	69,57%	30,43%
90%-10%	50%	50%

3. *Stratified Sampling*

Tabel 3. Hasil *Stratified Sampling*

Training Data-Testing Data	Accuracy	Classification Error
60%-40%	76,09%	23,91%
70%-30%	71,43%	28,57%
80%-20%	69,57%	30,43%
90%-10%	81,82%	18,18%

4. *Automatic Sampling*

Tabel 4. Hasil *Automatic Sampling*

Training Data-Testing Data	Accuracy	Classification Error
60%-40%	76,09%	23,91%
70%-30%	71,43%	28,57%
80%-20%	69,57%	30,43%
90%-10%	81,82%	18,18%

Pengujian dengan menggunakan operator *cross validation* dilakukan beberapa kali pengujian pada sampling datanya. Pengujian dilakukan agar dapat memilih tingkat akurasi yang paling bagus, berikut tabel pengujiannya:

Tabel 5. Tabel *Cross Validation*

Training Data-Testing Data	Accuracy	Classification Error
<i>Linear Sampling</i>	68,18%	31,82%
<i>Shuffled Sampling</i>	67,27%	32,73%
<i>Stratified Sampling</i>	66,52%	33,48%
<i>Automatic Sampling</i>	66,52%	33,48%

Berikut tabel prediksi hasil pengujian dengan menggunakan split validation dengan menggunakan data training 90% dan data testing 10% dengan type Stratified Sampling

Tabel 6. Tabel Prediksi

	True Tidak	True Ya	Class Precision
Prediksi Tidak	6	1	85,71%
Prediksi Ya	1	3	75,00%
Class Recall	85,71%	75,00%	

Setelah pengujian dilakukan beberapa kali maka operator yang ditetapkan adalah dengan menggunakan operator *split validation* dan menggunakan *Stratified Sampling* dengan training data 90% dan testing data 10% dengan tingkat akurasi sebesar 81,82% dan classification error sebesar 18,18%.

Tree

```

C01 = Baik
| C03 = Baik: Tidak {Tidak=6, Ya=3}
| C03 = Buruk: Ya {Tidak=0, Ya=2}
| C03 = Sangat Baik: Ya {Tidak=4, Ya=6}
| C03 = Sangat Buruk
| | C02 = Baik: Tidak {Tidak=1, Ya=1}
| | C02 = Sangat Baik: Ya {Tidak=0, Ya=2}
C01 = Buruk
| C03 = Baik
| | C04 = Baik: Ya {Tidak=1, Ya=3}
| | C04 = Buruk: Tidak {Tidak=3, Ya=2}
| | C04 = Sangat Baik: Ya {Tidak=0, Ya=4}
| C03 = Buruk: Tidak {Tidak=3, Ya=3}
| C03 = Sangat Baik: Tidak {Tidak=4, Ya=1}
| C03 = Sangat Buruk: Tidak {Tidak=3, Ya=0}
C01 = Sangat Baik
| C02 = Baik: Tidak {Tidak=5, Ya=2}
| C02 = Buruk: Tidak {Tidak=13, Ya=2}
| C02 = Sangat Baik
| | C05 = Baik: Ya {Tidak=1, Ya=3}
| | C05 = Buruk: Ya {Tidak=0, Ya=2}
| | C05 = Sangat Baik: Ya {Tidak=0, Ya=4}
| | C05 = Sangat Buruk: Tidak {Tidak=2, Ya=0}
| C02 = Sangat Buruk: Ya {Tidak=2, Ya=4}
C01 = Sangat Buruk: Tidak {Tidak=23, Ya=1}
    
```

Gambar 1. Rule Decision Tree

Berdasarkan gambar 1 diatas bisa dilihat bahwa C01(Seleksi Berkas)akan dijadikan node akar untuk mengetahui prediksi calon karyawan baru. Dengan tingkat akurasi yang diperoleh sebesar 81,82% dan classification error sebesar 18,18%.

PEMBAHASAN

Setelah dilakukan perhitungan manual dengan perhitungan di rapidminer terhadap calon karyawan baru menghasilkan 2 class yaitu class ya dan class tidak. Perhitungan manual yaitu jumlah class ya sebanyak 45 sedangkan class tidak sebesar 71. Sehingga dilihat bahwa calon karyawan yang tidak lolos lebih banyak daripada calon karyawan yang lolos.

KESIMPULAN

Berdasarkan penelitian tentang Penerapan Algoritma C4.5 Dalam Klasifikasi Penerimaan Calon Karyawan Biro Kesejahteraan Rakyat Berbasis Website, maka dapat diambil kesimpulan sebagai berikut:

1. Dataset yang digunakan merupakan data calon karyawan baru. Terdapat 6 atribut yang digunakan yaitu C01 (Seleksi Berkas), C02 (Test Buta Warna), C03 (Test Pengetahuan Sesuai Bidang), C04 (Efisiensi), C05 (Kerjasama), Rekomendasi sebagai target.
2. Pengujian dilakukan dengan menggunakan software rapidminer. Pengujian dilakukan dengan menggunakan 2 operator yang berbeda yaitu *split validation* dan *cross validation* dengan melakukan beberapa percobaan pada sampling data agar menemukan tingkat akurasi yang paling baik.
3. Setelah pengujian dilakukan beberapa kali maka operator yang ditetapkan adalah dengan menggunakan operator *split validation* dan menggunakan *stratified sampling* dengan *training data* 90% dan *testing data* 10% dengan tingkat akurasi yang diperoleh sebesar 81,82% dan *classification error* sebesar 18,18%. Dari hasil pengujian di rapidminer juga menghasilkan sembilan belas rule yang diperoleh untuk memprediksi calon karyawan baru yang akan diterima atau tidak.

REFERENSI

- Sukma, Halfis, dan Hermawan. (2019). Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4.5, Volume V No. 1 Februari 2019 P-ISSN 2442-2436, E-ISSN: 2550-0120
- Putri, Ratna Puspita Sari dan Waspada, Indra. Juni 2018. "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatik". Khazanah Informatika, Vol. 4 No. 1, Juni, 2019.
- Kusrini&Lutfi, E.T., 2019, Algoritma Data Mining, Penerbit Andi, Yogyakarta.
- Purnomo, Joko dkk. 2018. Implementasi algoritma C4.5 dalam pembuatan aplikasi penunjang keputusan penerimaan pegawai CV Dinamika Ilmu
- Merdeka.com.2021." Survei: Jumlah Lamaran per Lowongan Kerja Naik 89 Persen Selama Pandemi", <https://www.merdeka.com/uang/survei-jumlah-lamaran-per-lowongan-kerja-naik-89-persen-selama-pandemi.html>, diakses pada 31 Maret 2022.
- Moertini, V.S., 2020, Pengembangan Skalabilitas Algoritma Klasifikasi C4.5 Dengan Pendekatan Konsep Operator Relasi, Studi Kasus: Pra-pengolahan dan Klasifikasi Citra Batik, Disertasi, Program Studi Teknik Informatika, Institut Teknologi Bandung. Bandung.
- Nugroho, Y. S. (2019). Data mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas
- Ridwan, M., Suyono, H., & Sarosa, M. (2020). Penerapan Data mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, 7(1).
- Rohmawati, F., Rohman, G., & Mujilawati, S. (2020). Sistem Prediksi Jumlah Pengunjung Wisata Wego Kec.Sugio Kab.Lamongan Menggunakan Metode Fuzzy Time Series, 3(2)
- Witten, I. H., & Frank, E. (2019). Data mining: Practical Machine Learning Tools and Techniques. Elsevier, San Francisco, 3 edition, 629. <https://doi.org/0120884070>, 9780120884070.
- Hasibuan, D.Z. 2019. "Aplikasi Data Mining Dengan K-Means Cluster Untuk Memprediksi Produk Potensial Dan Penentuan Persediaan Produk". *Kumpulan Karya Ilmiah Mahasiswa Fakultas sains dan Tekhnologi* 1(1), 1- 61.
- Rahmaningsih, R. 2018. "Ekstraksi Informasi Postingan Jadwal Pengajian Pada Media Sosial Facebook Menggunakan Naive Bayes Classifier". *Seminar Nasional Teknologi Informasi Komunikasi dan Industri*.
- Maricar, M.A., Pramana, D. 2019. "Perbandingan Akurasi Naïve Bayes dan K- Nearest Neighbor Pada Klasifikasi Untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali". *Jurnal Sistem dan Informatika (JSI)* 14(1), 16- 22.
- Ardiyansyah, Rahayuningsih, P.A., Maulana, R. 2018. "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapidminer". *Jurnal Khatulistiwa Informatika* 6(1), 20-28.