

Perbandingan Metode Algoritma Naïve Bayes dan C4.5 dalam Menentukan Penilaian Kinerja Karyawan Terbaik

Suherman^{1*)}, Irfan Afriantoro²⁾, Slamet Mujiono³⁾

¹⁾²⁾³⁾ Fakultas Teknik, Universitas Pelita Bangsa

Correspondence author: suherman@pelitabangsa.ac.id, Cikarang, Indonesia

DOI: <https://doi.org/10.37012/jtik.v9i1.1605>

Abstrak

Karyawan terbaik merupakan salah satu asset perusahaan dan menjadi tolak ukur untuk kemajuan perusahaan itu sendiri. Dalam penilaian biasanya dengan mengevaluasi kinerja karyawan tersebut misalnya dari aspek kerajinan, kedisiplinan dan dari aspek prestasi lainnya. Penentuan karyawan terbaik sebelumnya masih kurang efektif. Hasil penilaian karyawan akan dimanfaatkan dan dievaluasi oleh pihak manajemen sebagai pengelola perusahaan. Banyaknya parameter dalam menentukan penilaian kinerja karyawan terbaik, menyebabkan ketepatan dan kecepatan dalam penilaian kerja kurang terpenuhi. Metode klasifikasi *data mining* merupakan sebuah teknik yang dilakukan untuk menghitung class atau label, menghitung jumlah kasus yang sama dengan class yang sama, mengalikan semua hasil variabel, membandingkan hasil class dengan hasil terbesar yang akan dijadikan pertimbangan untuk memutuskan karyawan terbaik. Algoritma klasifikasi yang sering digunakan dan mendapat banyak perhatian para peneliti dalam menentukan penilaian kinerja karyawan adalah algoritma Naïve Bayes dan algoritma C4.5. Perbandingan *accuracy* dari kedua algoritma dan hasil tertinggi akan menjadi hasil akhir. Dari 180 dataset dibagi menjadi 80% (144 data) training dan 20% (36 data) testing yang memiliki nilai *accuracy* tertinggi yaitu 85.07%. Hal ini menunjukkan bahwa algoritma Naive Bayes memiliki performa yang sangat baik dalam menentukan penilaian kinerja karyawan terbaik sehingga dapat di implemetasikan untuk proses pengambilan keputusan bagi perusahaan.

Kata kunci: Naïve Bayes, Algoritma C4.5, Kinerja Karyawan

ABSTRACT

The best employees are one of the company's assets and a benchmark for the progress of the company itself. In the assessment usually by evaluating the employee's performance, for example from aspects of diligence, discipline and from other aspects of achievement. Determination of the best employees before is still less effective. The results of the employee assessment will be utilized and evaluated by the management as the manager of the company. The number of parameters in determining the best employee performance appraisal, causes the accuracy and speed in evaluating work to be less fulfilled. The data mining classification method is a technique that is used to calculate classes or labels, calculate the number of cases that are the same with the same class, multiply all variable results, compare class results with the largest results which will be taken into consideration to decide on the best employee. The classification algorithms that are often used and get a lot of attention from researchers in determining employee performance appraisals are the Naïve Bayes algorithm and the C4.5 algorithm. Comparison of the accuracy of the two algorithms and the highest result will be the final result. Of the 180 datasets divided into 80% (144 data) training and 20% (36 data) testing which has the highest accuracy value of 85.07%. This shows that the Naive Bayes algorithm has a very good performance in determining the best employee performance appraisal so that it can be implemented for the company's decision-making process.

Keywords: Naive Bayes, C4.5 Algorithm, Employee Performance

PENDAHULUAN

Karyawan terbaik merupakan salah satu asset perusahaan dan menjadi tolak ukur untuk kemajuan perusahaan itu sendiri. Dalam penilaian biasanya dilakukan dengan mengevaluasi kinerja karyawan tersebut misalnya dari aspek kerajinan, kedisiplinan dan dari aspek prestasi lainnya. Penentuan karyawan terbaik sebelumnya masih kurang efektif. Hasil penilaian ini akan dimanfaatkan dan dievaluasi oleh pihak manajemen sebagai pengelola perusahaan. Manipulasi data dapat terjadi karena sulitnya mengambil keputusan dengan kriteria dan data yang begitu besar tanpa adanya otomatisasi penggalan data. Sebagai dampaknya perusahaan akan sulit mendapatkan sumber daya manusia yang kompetitif. Salah satu caranya yaitu dengan mengembangkan metode penelitian baru yang dapat menambah keakuratan hasil keputusan yaitu penggunaan *data mining*.

Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna bagi perusahaan. Banyaknya parameter dalam menentukan penilaian kinerja karyawan terbaik, menyebabkan ketepatan dan kecepatan dalam penilaian kerja kurang terpenuhi. Metode klasifikasi data mining merupakan sebuah teknik yang dilakukan untuk memprediksi class atau label dari data itu sendiri. Beberapa Algoritma yang sering digunakan dalam memprediksi masalah penilaian kinerja karyawan terbaik di sebuah Perusahaan adalah Algoritma Naïve Bayes dan Algoritma C4.5. Kesederhanaan pada Algoritma Naïve Bayes dan Algoritma C4.5 yang membuat kedua Algoritma tersebut mempunyai daya tarik untuk diimplementasikan dalam berbagai sistem. Adapun kelemahan yang ada pada kedua Algoritma tersebut adalah lamanya waktu dan tingkat akurasi prediksi yang digunakan untuk melakukan prediksi. Masalah ini juga menjadi perhatian banyak peneliti *Machine learning* atau pembelajaran mesin. *Machine Learning* adalah ilmu atau studi yang mempelajari tentang algoritma dan model statistic yang digunakan oleh sistem komputer untuk melakukan task tertentu tanpa instruksi eksplisit. *Machine learning* bergantung pada pola dan kesimpulan. Untuk mendapatkan pola dan kesimpulan tersebut, algoritma *machine learning* menghasilkan model matematika yang didasari dari data sampel yang sering disebut dengan “training data” (Hania, A.A. and Y.T).

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 data base besar. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Data mining, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam data set berukuran besar (Ridwan, M., Suyono, H., and Sarosa M.).

Prediksi adalah menentukan jumlah kebutuhan bulan mendatang terkait dengan dukungan data historis (*historical data*) atau serangkaian waktu / periode yang dianalisis sehingga dapat diperhitungkan untuk memprediksi jumlah kebutuhan pada bulan mendatang. Prediksi juga dapat digunakan dalam pengklasifikasian, tidak hanya untuk memprediksi *time series*, karena sifatnya yang bisa menghasilkan class berdasarkan atribut yang ada.

METODE

Dalam melakukan penelitian untuk mendapatkan data dan informasi, maka metode yang digunakan dalam proses pengumpulan data yaitu:

1. Pada penelitian dimulai dengan mempelajari informasi serta algoritma yang bersangkutan dengan penelitian ini dengan cara mendownload serta membaca ebook, ejournal, dan beberapa referensi pembelajaran lain. Pada tahapan ini konsep-konsep yang dibutuhkan dalam penelitian akan dimatangkan, seperti definisi dari data mining, Algoritma Naive Bayes dan Algoritma C4.5 serta penerapannya.
2. Observasi, metode yang digunakan untuk menganalisa dan mengamati secara langsung serta untuk mendapatkan informasi terhadap objek penelitian maupun data penelitian.

Sumber data adalah data yang didapatkan dari objek penelitian secara langsung dan akan digunakan sebagai bahan penelitian. Data yang akan di teliti dan di proses ini sudah dipertimbangkan dari segi jenis data maupun keakuratan data, guna mendukung indikator keberhasilan dalam penelitian.

Data yang digunakan pada penelitian ini meliputi:

1. Data Primer

Data yang diperoleh atau dikumpulkan oleh peneliti secara langsung dari objek penelitian, mengenai hal-hal yang berhubungan dengan penulisan skripsi ini.

2. Data Sekunder

Data yang diperoleh secara tidak langsung berasal dari buku-buku referensi atau literatur lainnya yang berhubungan dengan penulisan penelitian ini, meliputi pendapat dan penjelasan yang dikemukakan oleh para ahli maupun peneliti-peneliti sebelumnya sehingga membantu dalam penelitian.

Data asli yang sangat banyak tidak bisa langsung di proses dengan *data mining*. Maka tahap pengelompokan data digunakan untuk mengelompokkan data yang diperlukan. Data dari penilaian kinerja akan di lakukan *cleaning data*, kemudian dilakukan *data selection* yang akan menghasilkan *transformed data*.

Seleksi data menjadi hal yang sangat penting dalam seleksi data maka perlu pengenalan terhadap data yang akan diolah. Tidak semua data penelitian sudah memiliki atribut dan variabel sendiri sebelum dilakukan proses *mining*. Maka seleksi data atau mengklasifikasi data dilakukan untuk menentukan atribut dan variabel yang dibutuhkan untuk mengolah data.

Hasil dari proses seleksi 180 data yang akan digunakan menghasilkan 20 atribut. Atribut berisi masing-masing nilai yang sudah ditentukan, hal ini akan menentukan hasil ataupun akurasi yang akan dihasilkan. Adapun variabel nilai yang akan digunakan dalam proses penelitian ini dapat dilihat pada tabel 1.

Tabel 1. Kriteria Variabel Penilaian

Variabel Nilai	
Variabel	Nilai
Kurang Sekali	1
Kurang	2
Cukup	3
Baik	4
Baik Sekali	5

Aspek penilaian karyawan dalam penelitian ini terdiri dari 20 aspek penilaian, yang diberi kode A1 sampai A20 sebagaimana terlihat pada tabel 2 berikut ini.

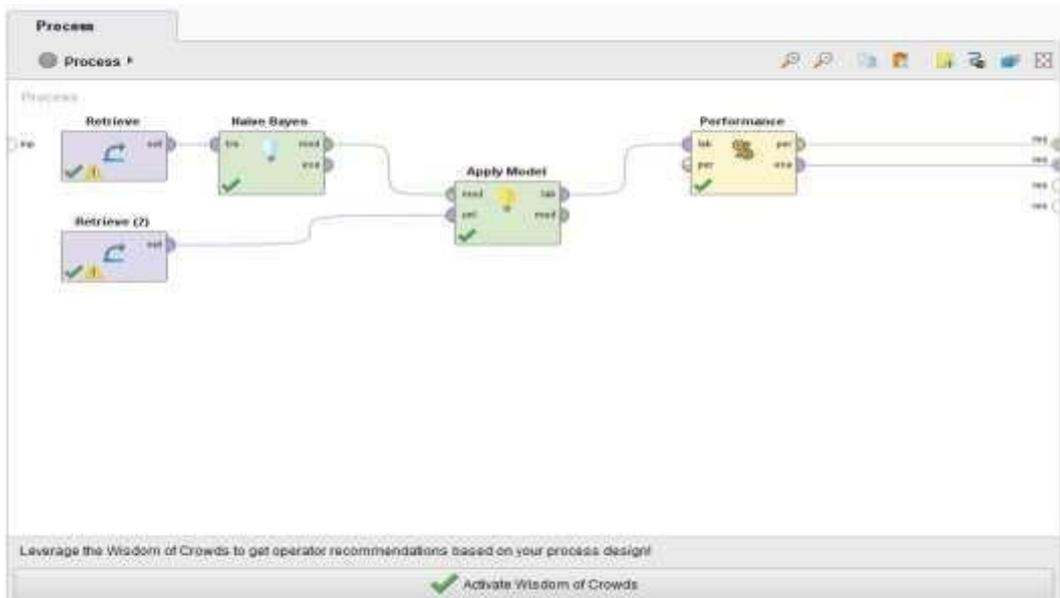
Tabel 2. Aspek Penilaian Karyawan

Aspek Penilaian	
A1	Kemampuan mencapai target produksi / kerja sesuai kualitas yang sudah ditentukan
A2	Kemampuan menetapkan skala prioritas pekerjaan
A3	Kemampuan membuat keputusan secara cepat dan tepat
A4	Pengetahuan yang dimiliki sesuai lingkup pekerjaannya
A5	Kemampuan bekerja sama dan berkomunikasi dengan Pimpinan atau TeamWork
A6	Kemampuan melakukan kontrol pengeluaran dan melakukan penghematan atau Efisiensi
A7	Kreatifitas yang dimiliki untuk mengatasi persoalan baru dalam pekerjaan
A8	Kepedulian terhadap lingkungan disekitarnya
A9	Kemampuan peng-operasian alat kerja sehari-hari
A10	Loyalitas dalam bekerja
A11	Kedisiplinan dalam bekerja
A12	Kehadiran setiap hari (Absensi)
A13	Kelengkapan Seragam Kerja setiap hari
A14	Efisiensi dalam penggunaan waktu jam kerja
A15	Kemampuan menyesuaikan diri dengan lingkungan dengan situasi yang berbeda
A16	Memiliki pengaruh yang positif dalam lingkungan kerja
A17	Kemampuan mempengaruhi rekan kerja untuk berprestasi secara individual ataupun kelompok
A18	Inisiatif yang dimiliki pada saat diperlukan
A19	Keinginan menerima tantangan / tanggung jawab yang lebih besar
A20	Kemauan untuk belajar guna mengembangkan diri di masa depan

Data yang siap proses berjumlah 180 data, 20 atribut dan 2 label kelas transformasi data ini akan digunakan sebagai *training data* dan *testing data* yang akan digunakan pada proses pengujian.

HASIL DAN PEMBAHASAN

Berikut adalah proses pengujian menggunakan Naïve Bayes.



Gambar 1. Proses Pengujian Naïve Bayes

Tabel 3 dan 4 berikut ini adalah proses perhitungan *Accuracy* dan *Precision* menggunakan Naïve Bayes.

Tabel 3. Hasil *Accuracy* Naïve Bayes

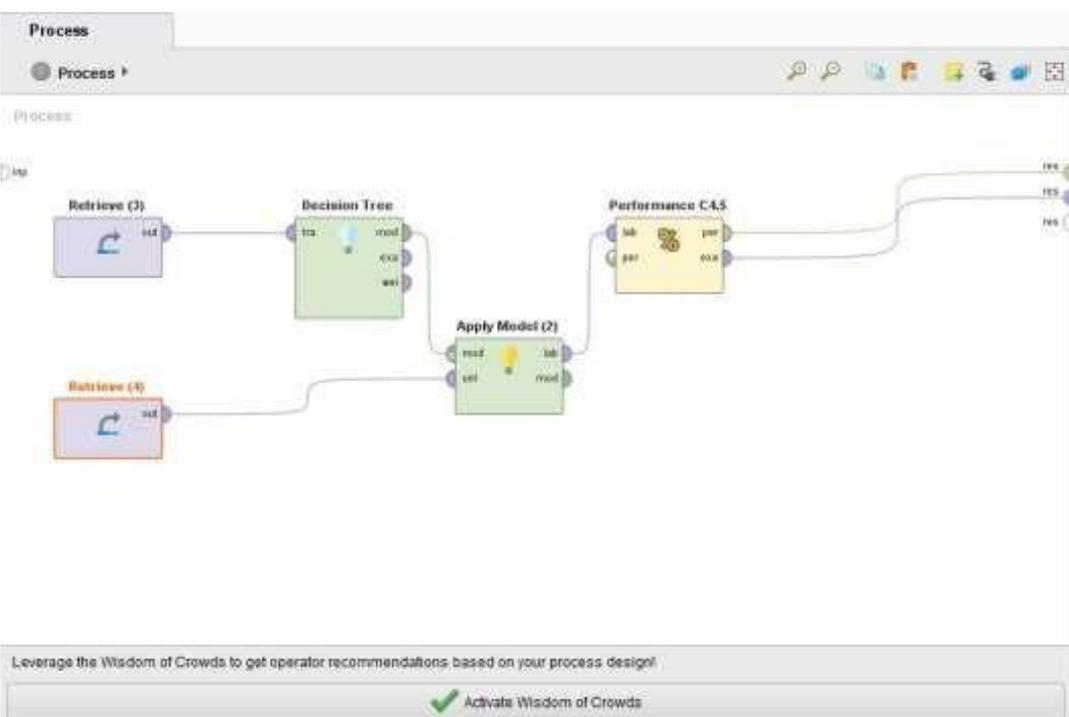
accuracy: 86.11%

	true TIDAK BAIK	true BAIK	class precision
pred TIDAK BAIK	15	4	78.95%
pred BAIK	1	16	94.12%
class recall	93.75%	80.00%	

Tabel 4. Hasil *Precision* Naïve Bayes

precision: 94.12% (positive class: BAIK)

	true TIDAK BAIK	true BAIK	class precision
pred. TIDAK BAIK	15	4	78.95%
pred. BAIK	1	16	94.12%
class recall	93.75%	80.00%	



Gambar 2. Proses Pengujian Algoritma C4.5

Berdasar dari hasil *confusion matrix*, selanjutnya dilakukan perhitungan nilai *accuracy*, *precision* dan *recall*. Perbandingan nilai *accuracy*, *precision* dan *recall* yang telah dihitung dengan metode naive bayes dan C4.5 sebagai berikut:

Tabel 5. Perbandingan nilai *accuracy*, *precision* dan *recall* metode Naive bayes dan C4.5

	Naïve Bayes		C4.5	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
<i>Accuracy</i>	84.03%	86.11%	75.00%	88.89%
<i>Precision</i>	78.35%	94.12%	71.88%	86.36%
<i>Recall</i>	97.44%	80.00%	88.46%	95.00%

Tabel 6 berikut ini menyatakan perbandingan hasil perhitungan nilai AUC untuk metode Naïve Bayes dan C.45.

Tabel 6. Nilai AUC Naïve Bayes dan C4.5

	Naïve Bayes		C4.5	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
<i>AUC</i>	0.920	0.919	0.554	0.889

Tabel 7. *Confusion Matrix* metode Naive bayes dan C4.5

Metode	<i>Confusion Matrix Accuracy</i>		Perbandingan
	<i>Training</i>	<i>Testing</i>	
Naïve Bayes	84.03%	86.11%	85.07%
C4.5	75.00%	88.89%	81.95%

Berdasarkan hasil perbandingan akurasi, Algoritma Naive Bayes memiliki tingkat akurasi lebih baik sehingga bisa digunakan untuk menentukan penilaian kinerja karyawan terbaik dengan persentase 85.07%.

KESIMPULAN DAN REKOMENDASI

Berdasarkan penelitian yang dilakukan, dapat ditarik kesimpulan sebagai berikut, bahwa Perusahaan mendapatkan sumber daya manusia yang kompetitif dengan adanya standar penilaian kinerja karyawan, Perusahaan lebih mudah dalam melakukan penilaian kinerja karyawan dengan metode yang digunakan. Akurasi dari pengujian dengan *tools* Rapidminer memudahkan proses implementasi dalam melakukan penilaian kinerja karyawan.

Proses pengujian menunjukkan bahwa algoritma Naive Bayes menghasilkan tingkat akurasi tertinggi sebesar 85.07%, sedangkan algoritma C4.5 menunjukkan akurasi sebesar 81.95%. Target tingkat akurasi dari penelitian perbandingan metode algoritma naïve bayes dan C4.5 dalam menentukan penilaian kinerja karyawan terbaik yaitu 80% dari yang sebelumnya hanya 75%. Tingkat akurasi mencapai peningkatan sebesar 10.07% dari penilaian sebelumnya dan target mencapai peningkatan sebesar 5.07%. Hal ini menunjukkan bahwa pengujian terbaik dari kedua algoritma yang telah dilakukan pada penelitian ini adalah algoritma Naive Bayes dengan selisih dari pengujian algoritma C4.5 sebesar 3.13%. Dengan demikian bahwa algoritma Naive Bayes merupakan algoritma terbaik dalam menentukan penilaian kinerja karyawan terbaik.

REFERENSI

- Rosandy, T. (2016). “Perbandingan Metode Naïve Bayes Classifier Dengan Metode Decision Tree (C4.5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus Kspps Bmt Al Fadhila)”, vol.02, no.01, pp.52–62.
- Bahri, S., Midyanti, D. M. and Hidayati, R. (2018), “Perbandingan Algoritma Naïve Bayes dan C4.5 Untuk Klasifikasi Penyakit Anak”, pp.24–31.
- Julianto, W., Yunitaraini, R. and Sophan, (2014). “Algoritma C4.5 Untuk Penilaian Kinerja Karyawan,” vol.IX, pp.33–39.
- Sunge, A. S., et al., (2018). “Prediksi Kompetensi Karyawan Menggunakan Algoritma C4.5 (Studi Kasus : PT Hankook Tire Indonesia)”, Sentika, pp.23–24.

-
- Ridwan, M., Suyono, H., and Sarosa, M., (2013). “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier,” vol.7, no.1, pp.59–64.
- Faisal, A., (2017). “Network Untuk Memenuhi Penilaian Data Karyawan Service Level Agreement Di Bank” vol.10, no.4, pp.350–361.
- Hania, A. A., and Y.T., (2017). “Mengenal Artificial Intelligence, Machine Learning, Neural Network dan Deep Learning,” June.
- Ma’ruf Abdulah, S. M., (2014). Manajemen dan Evaluasi Kinerja Karyawan.
- Ahmad, S., (2018). “Perbandingan Partisipasi Politik Masyarakat dalam Pemilihan Kepala Daerah (Pilkada) Serentak Tahun 2015”, vol. 6, no. 2, pp. 171–181.
- Saleh, A. (2015). “Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,” vol. 2, no. 3, pp. 207–217.
- Sari, V. N., Astri, L. Y., Rasywir, E., (2020). “Analisis dan Penerapan Algoritma Naive Bayes Untuk Evaluasi,” vol. 2, no. 1, pp. 53–68.
- Masripah, S., (2016)., “Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit”, vol.3, no.1, pp.187–193.