



Data Mining Menggunakan Algoritma Naïve Bayes Classifier Untuk Evaluasi Kinerja Karyawan

Bunga Haya Pangestu*

Prodi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ARTICLE INFO

Article history :

Received : 20/9/2023
Revised : 1/12/2023
Published : 12/12/2023



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Volume : 3
No. : 2
Halaman : 177-184
Terbitan : **Desember 2023**

ABSTRAK

Data Mining adalah proses dalam menemukan pola tersembunyi, hubungan, atau informasi yang berharga dari data yang besar, kompleks, dan terstruktur. Penelitian ini bertujuan untuk membahas mengenai Algoritma *Naïve Bayes Classifier* dalam proses *Data Mining* untuk mengevaluasi kinerja karyawan sehingga dapat ditentukan perpanjangan kontrak kerja karyawan. Penelitian ini berdasarkan kriteria kualitas pekerjaan, kuantitas pekerjaan, analisis/pemecahan masalah, daya tangkap, amanah, kompeten, harmonis, loyal, adaptif, dan kolaboratif. Algoritma *Naïve Bayes Classifier* digunakan untuk memprediksi kinerja karyawan berdasarkan atribut. Penelitian ini menghasilkan *knowledge* yaitu nilai *accuracy* sebagai tingkat kedekatan antara nilai prediksi dengan nilai fakta, *precision* sebesar sebagai ketepatan antara informasi yang ada dengan hasil yang diperoleh, *recall* sebagai tingkat keberhasilan model dalam menemukan kembali sebuah informasi, dan *error rate* sebagai tingkat kekeliruan.

Kata Kunci : *Data mining; Naïve Bayes Classifier; Evaluasi Kinerja karyawan.*

ABSTRACT

Data Mining is the process of finding hidden patterns, relationships or valuable information from large, complex and structured data. This study aims to discuss the *Naïve Bayes Classifier* Algorithm in the Data Mining process to evaluate employee performance so that employee contract extensions can be determined. This research is based on the criteria of quality of work, quantity of work, analysis/problem solving, comprehension, trustworthiness, competence, harmony, loyalty, adaptive, and collaborative. The *Naïve Bayes Classifier* algorithm is used to predict employee performance based on attributes. This study produces knowledge, namely the accuracy value as the degree of closeness between the predicted value and the fact value, precision as the accuracy between the existing information and the results obtained, recall as the success rate of the model in retrieving an information, and error rate as the level of error.

Keywords : *Data mining; Naïve Bayes Classifier; Employee Performance Evaluation.*

@ 2023 Jurnal Riset Matematika, Unisba Press. All rights reserved.

A. Pendahuluan

Setiap perusahaan membutuhkan adanya karyawan untuk menjalankan fungsi-fungsi yang terkandung di dalam perusahaan. Karyawan memiliki dampak yang sangat besar dalam keberhasilan suatu perusahaan. Karyawan menjadi salah satu sumber daya utama dalam perusahaan untuk memperlancar produktivitas. Maka dari itu, karyawan dituntut untuk selalu memberikan kontribusi yang efektif dan efisien dalam bidangnya. Untuk membentuk suatu kompetensi kinerja karyawan tersebut dapat diberikan suatu evaluasi kinerja agar dapat dilakukan penilaian secara sistematis dalam mengetahui hasil kinerja karyawan dalam berkontribusi pada perusahaan. Konsep evaluasi dapat disamakan dengan penaksiran (*appraisal*), pemberian angka (*rating*) dan penilaian (*assessment*). Evaluasi kinerja sangat penting dalam mengevaluasi tanggung jawab seseorang dalam organisasi [1][2].

Data Mining akan digunakan untuk menentukan kelayakan karyawan tertentu [3][4]. Algoritma *Naive Bayes Classifier* [5]. *Knowledge Discovery in Database* (KDD) adalah penerapan metode saintifik pada data mining [6]. Dalam konteks ini, data mining merupakan satu langkah dari proses KDD. Data Mining merupakan salah satu proses, maka dipecah beberapa tahapan yaitu koleksi data, seleksi data, preprocessing, transformasi data, Data Mining yang akan digunakan adalah Algoritma *Naive Bayes Classifier*, Evaluasi, KDD [7][8]. *Naive Bayes* adalah suatu klasifikasi secara sederhana di mana untuk menghitung probabilitas dengan cara menjumlahkan frekuensi dan nilai dataset aslinya [9][10]. Data Mining merupakan salah satu bagian tahapan dari *Knowledge Discovery in Database* [11][12].

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “ Bagaimana proses *Data Mining* mengevaluasi kinerja karyawan menggunakan Algoritma *Naive Bayes Classifier*?”. “Bagaimana akurasi model *Naive Bayes Classifier* dalam menentukan status karyawan”. Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb. (1) Mengetahui implementasi Data Mining untuk evaluasi kinerja karyawan; (2) Mengetahui identifikasi dan klasifikasi evaluasi karyawan menggunakan metode *Naive Bayes Classifier*.

B. Metode Penelitian

Peneliti menggunakan teknik data mining menggunakan Algoritma *Naive Bayes Classifier*. Data yang dipilih dalam penelitian ini adalah karyawan anak perusahaan BUMN XYZ yang berjumlah 112 karyawan.

Teknik pengumpulan data yang digunakan penelitian ini adalah wawancara, observasi, dan studi Pustaka. Dengan melakukan tahapan *data mining* dalam penelitian ini diperoleh jumlah data penelitian sebanyak 100 data karyawan.

Pada tahap ini bertujuan untuk memprediksi apakah *Naive Bayes classifier* dapat mengklasifikasikan data penelitian ini. Berikut proses *Naive Bayes classifier*.

Data training sesuai variabel class “YA” dan “TIDAK”. *Data training* berjumlah 70 data, kemudian dikelompokkan berdasarkan variabel *class*.

Input Nilai. Input nilai kedalam tabel pengujian klasifikasi sesuai nilai peluang kriteria berdasarkan variabel *class* “YA” dan “TIDAK” yang sudah ada.

Menghitung peluang kriteria. Berdasarkan kriteria yang sudah ada dengan berbagai variabel nilai, kemudian menghitung peluang masing-masing kriteria dengan cara memasukkan variabel nilai kedalam data yang sudah dikelompokkan.

Hasil klasifikasi dengan status aslinya. Pada tahap ini akan menghasilkan hasil klasifikasi yang bertujuan apakah karyawan tersebut layak atau tidaknya karyawan tersebut diperpanjang berdasarkan klasifikasi *Naive Bayes classifier*.

C. Hasil dan Pembahasan

Proses Data Mining

Tahapan data mining untuk menghasilkan prediksi klasifikasi evaluasi kinerja karyawan menggunakan klasifikasi *Naive Bayes classifier*. Klasifikasi *Naive Bayes classifier* terdapat koleksi data, *preprocessing*, klasifikasi *Naive Bayes classifier*, dan uji akurasi.

Preprocessing

Preprocessing dilakukan untuk pengelompokkan data training dan data testing. Data yang didapatkan yaitu 112 data dengan menggunakan variabel nilai 1 sampai dengan 5. Pada tahap preprocessing terdapat koleksi data dan seleksi data.

Koleksi Data. Koleksi data diambil sebanyak 112 data karyawan berupa evaluasi kinerja karyawan dengan jangka waktu tahun 2022 sampai dengan 2023 dengan menggunakan beberapa atribut variabel penelitian.

Seleksi Data. Seleksi Data dilakukan karena tidak semua data dapat digunakan dalam proses *mining*. Data yang digunakan harus sudah mencakup informasi yang dijadikan indikator penelitian.

Input Data

Input Data dilakukan dari hasil dari pengelompokkan *preprocessing* yang akan diolah ke dalam bentuk *Microsoft Excel*.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Kary.	A	B	C	D	E	F	G	H	I	J	Status
2	A1	Cukup	Baik	Cukup	Cukup	Baik	Cukup	Baik	Baik	Baik	Baik	YA
3	A2	Baik	Cukup	Cukup	Baik	Sangat Baik	Baik	Baik	Baik	Cukup	Cukup	YA
4	A3	Baik	Cukup	Cukup	Cukup	Baik	Cukup	Baik	Baik	Baik	Baik	TIDAK
5	A4	Baik	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Baik	Sangat Baik	TIDAK
6	A5	Cukup	Cukup	Cukup	Cukup	Sangat Baik	Baik	Baik	Baik	Baik	Baik	TIDAK
7	A6	Baik	Baik	Baik	Baik	Sangat Baik	Baik	Baik	Baik	Baik	Baik	TIDAK
8	A7	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Baik	Baik	Baik	Baik	Baik	YA
9	A8	Sangat Baik	Sangat Baik	Baik	Sangat Baik	Baik	Sangat Baik	Baik	Sangat Baik	Sangat Baik	Sangat Baik	YA
10	A9	Baik	Baik	Baik	Baik	Sangat Baik	Baik	Baik	Baik	Sangat Baik	Sangat Baik	YA
11	A10	Baik	Baik	Baik	Baik	Sangat Baik	Baik	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	YA
12	A11	Baik	Sangat Baik	Baik	Sangat Baik	Sangat Baik	Baik	Sangat Baik	Sangat Baik	Baik	Sangat Baik	TIDAK
13	A12	Sangat Baik	Sangat Baik	Cukup	Baik	Sangat Baik	Baik	Sangat Baik	Sangat Baik	Baik	Sangat Baik	TIDAK
14	A13	Cukup	Kurang	Cukup	Cukup	Cukup	Cukup	Baik	Baik	Cukup	Baik	TIDAK
15	A14	Baik	Cukup	Cukup	Baik	Baik	Cukup	Cukup	Baik	Cukup	Cukup	TIDAK

Gambar 1. Input Data.

Klasifikasi Naïve Bayes Classifier

Proses perhitungan untuk memprediksi apakah *Naïve Bayes classifier* dapat mengklasifikasikan data karyawan perusahaan XYZ. Berikut proses perhitungan *Naïve Bayes classifier*.

Menghitung Jumlah *Class*. Bentuk umum dari teorema bayes dapat dilihat dari persamaan berikut [4].

$$P(H | X) = \frac{P(X | H) \cdot P(H)}{P(X)} \tag{1}$$

Data *class* “YA” sebanyak 42 data karyawan dan data “TIDAK” sebanyak 28 data dengan menggunakan persamaan diatas.

	YA	TIDAK	
P[YA/TIDAK]	0.60	0.40	1

Gambar 2. Nilai Probabilitas *Class*/label.

Menghitung Jumlah Kasus yang Sama dengan Class yang Sama. Proses klasifikasi memerlukan suatu petunjuk untuk menentukan class apa yang cocok untuk dianalisis [4].

$$P(C | F1 \dots Fn) = \frac{P(C) \cdot P(F1 \dots Fn | C)}{P(F1 \dots Fn)} \tag{2}$$

Data yang digunakan pada perhitungan ini yaitu dengan mencari peluang di setiap kasus atau kriteria. Terdapat 10 kriteria dalam kasus data ini.

Tabel 1. Contoh Mencari Peluang Kriteria.

MENCARI PELUANG KRITERIA A				
A	YA	TIDAK	P [A YA]	P [A TIDAK]
Sangat Baik	14	5	0.33	0.18
Baik	24	17	0.57	0.61
Cukup	4	6	0.10	0.21
Total	42	28	1.00	1.00
Grand Total	70		2.00	

Menguji Hasil Klasifikasi *Naive Bayes Classifier*. Data yang akan digunakan sebagai pengujian hasil klasifikasi berjumlah 30 data testing dan yang akan dicari adalah status karyawan selanjutnya apakah karyawan tersebut layak atau tidak diperpanjang.

Kelompokkan data training sesuai variabel class “YA” dan “TIDAK”.

Input nilai kedalam tabel pengujian dan kalikan seluruh kriteria A sampai dengan J sesuai variabel class “YA” dan “TIDAK”.

Menghitung peluang masing-masing kriteria tujuan klasifikasi ini yaitu mengetahui status karyawan selanjutnya. Apabila $P [YA] \geq P [TIDAK]$, maka hasil klasifikasi “YA” dan apabila $P [YA] < P [TIDAK]$, maka hasil klasifikasi “TIDAK”.

Tabel 2. Hasil Perkalian Seluruh Kriteria.

Karyawan	P [Karyawan YA]	P [Karyawan TIDAK]	Klasifikasi
K71	0.0000000000000000	0.0000004873179461	<i>TIDAK</i>
K72	0.00006826229015	0.0000001372147321	<i>YA</i>
K73	0.00006826229015	0.0000001372147321	<i>YA</i>
K74	0.0000000000000000	0.0000000886032629	<i>TIDAK</i>
K75	0.0000000000000000	0.0000000000000000	<i>YA</i>
K76	0.00006826229015	0.0000001372147321	<i>YA</i>

Karyawan	P [Karyawan YA]	P [Karyawan TIDAK]	Klasifikasi
K77	0.00000027922861	0.0000347945013546	<i>TIDAK</i>
K78	0.00037775365098	0.0000472517919630	<i>YA</i>
K79	0.00000050450000	0.0000026881019428	<i>TIDAK</i>
K80	0.00006826229015	0.0000001372147321	<i>YA</i>
K81	0.00000000000000	0.0000416515795822	<i>TIDAK</i>
K82	0.00009386064896	0.0000002156231505	<i>YA</i>
K83	0.00009386064896	0.0000002156231505	<i>YA</i>
K84	0.00051374496534	0.0000534528604638	<i>YA</i>
K85	0.00000000000000	0.0000015948587328	<i>TIDAK</i>
K86	0.00000005584572	0.0000347945013546	<i>TIDAK</i>
K87	0.00000005584572	0.0000347945013546	<i>TIDAK</i>
K88	0.00014197409086	0.0000037703248029	<i>YA</i>
K89	0.00003323518544	0.0002164991195395	<i>TIDAK</i>
K90	0.00016049222884	0.0000007470579860	<i>YA</i>
K91	0.00015050422142	0.0000012027728863	<i>YA</i>
K92	0.00003723571703	0.0001984575262445	<i>TIDAK</i>
K93	0.00019182036043	0.0000707464329668	<i>YA</i>
K94	0.00248238113504	0.0005154382973295	<i>YA</i>
K95	0.00248238113504	0.0005154382973295	<i>YA</i>
K96	0.00051831810948	0.0000259209830197	<i>YA</i>
K97	0.00000120130354	0.0001183404254073	<i>TIDAK</i>
K98	0.00013384762774	0.0000006468694515	<i>YA</i>
K99	0.00248238113504	0.0005154382973295	<i>YA</i>
K100	0.00000000194246	0.0000124266076266	<i>TIDAK</i>

Membandingkan Hasil Klasifikasi dengan status aslinya. Pada tahap ini dilakukan dengan cara membandingkan antara hasil klasifikasi dengan status aslinya yaitu kontrak maupun magang. Status karyawan ini tidak berpengaruh selama klasifikasi menunjukkan “YA”, maka diperpanjang kontrak dan apabila klasifikasi menunjukkan “TIDAK”, maka tidak diperpanjang. Tujuan pada tahap ini yaitu untuk mengetahui apakah karyawan tersebut layak atau tidak dilakukan perpanjangan kontrak atau tidak.

Tabel 3. Hasil Klasifikasi dengan Fakta.

Karyawan	Klasifikasi	Status Karyawan	KDD
K71	<i>TIDAK</i>	KONTRAK	<i>Tidak Lanjut Kontrak</i>
K72	<i>YA</i>	KONTRAK	<i>Lanjut Kontrak</i>
K73	<i>YA</i>	KONTRAK	<i>Lanjut Kontrak</i>
K74	<i>TIDAK</i>	MAGANG	<i>Tidak Lanjut Magang</i>
K75	<i>YA</i>	KONTRAK	<i>Tidak Lanjut Kontrak</i>
K76	<i>YA</i>	MAGANG	<i>lanjut Magang</i>

Karyawan	Klasifikasi	Status Karyawan	KDD
K77	TIDAK	KONTRAK	Tidak Lanjut Kontrak
K78	YA	KONTRAK	Lanjut Kontrak
K79	TIDAK	KONTRAK	Tidak Lanjut Kontrak
K80	YA	KONTRAK	Lanjut Kontrak
K81	TIDAK	MAGANG	Tidak Lanjut Magang
K82	YA	KONTRAK	Lanjut Kontrak
K83	YA	KONTRAK	Lanjut Kontrak
K84	YA	KONTRAK	Lanjut Kontrak
K85	TIDAK	KONTRAK	Tidak Lanjut Kontrak
K86	TIDAK	KONTRAK	Tidak Lanjut Kontrak
K87	TIDAK	KONTRAK	Tidak Lanjut Kontrak
K88	YA	KONTRAK	Lanjut Kontrak
K89	TIDAK	KONTRAK	Tidak Lanjut Kontrak
K90	YA	KONTRAK	Lanjut Kontrak
K91	YA	KONTRAK	Lanjut Kontrak
K92	TIDAK	KONTRAK	Tidak Lanjut Kontrak
K93	YA	KONTRAK	Lanjut Kontrak
K94	YA	KONTRAK	Lanjut Kontrak
K95	YA	KONTRAK	Lanjut Kontrak
K96	YA	KONTRAK	Lanjut Kontrak
K97	TIDAK	KONTRAK	Tidak Lanjut Kontrak
K98	YA	KONTRAK	Lanjut Kontrak
K99	YA	KONTRAK	Lanjut Kontrak
K100	TIDAK	KONTRAK	Tidak Lanjut Kontrak

Hasil Klasifikasi *Naive Bayes Classifier*. Setelah diperoleh hasil klasifikasi, terdapat 30 data testing. Dengan hasil “perpanjangan kontrak” berjumlah 17 karyawan dengan presentase yaitu 56.67% dari data testing dan “tidak perpanjangan kontrak” berjumlah 13 karyawan dengan presentase 43.33%.

Uji Akurasi

Uji akurasi merupakan proses pengukuran berdasarkan tingkat keakuratan prediksinya. Proses pengujiannya berdasarkan *metrics* atau tabel dari nilai prediksi dan nilai aktual. Namun, uji akurasi bukan satu-satunya akurasi yang dapat memberikan gambaran yang akurat tentang kinerja model [5].

Tabel 4. Bentuk *Metrics*.

		OBSERVED CLASS	
		KONTRAK	MAGANG
PREDICTED CLASS	YA	16	1
	TIDAK	11	2

Hasil uji akurasi menggunakan algoritma *Naïve Bayes classifier* dapat diukur menggunakan *accuracy*, *precision*, *recall*, dan *error rate* [5]. Pada akurasi klasifikasi *data mining*, nilai *Area Under Curve* (AUC) dapat dibagi menjadi beberapa kelompok, yaitu [6]:

1. 0,90 – 1,00 = Klasifikasi sangat baik
2. 0,80 – 0,90 = Klasifikasi baik
3. 0,70 – 0,80 = Klasifikasi cukup
4. 0,60 - 0,70 = Klasifikasi buruk

Tabel 5. Hasil Uji Akurasi *Metrics*.

<i>Accuracy</i>	60%
<i>Precision</i>	94%
<i>Recall</i>	59%
<i>Error rate</i>	10%

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut: (1) Penelitian ini menggunakan 100 data evaluasi kinerja karyawan. Kemudian, dilakukan tahapan KDD (*Knowledge Discovery in Database*) untuk mendapatkan berapa karyawan yang akan perpanjangan kontrak menggunakan Algoritma *Naïve Bayes Classifier*. Seluruh data yang digunakan dalam penelitian ini dibagi menjadi 2 (dua), yaitu 70% sebagai *data training* dan 30% sebagai *data testing*. 70% *data training* digunakan sebagai alat ukur untuk menghitung 30% *data testing*. *Knowledge* yang dihasilkan ini adalah banyaknya karyawan yang perpanjangan kontrak. Dengan 10 kriteria, banyaknya presentase karyawan yang diperpanjang yaitu 56.67%. Sedangkan, 43,33% yang tidak diperpanjang dari *data testing*. Kriteria analisa/pemecahan masalah, daya tangkap, dan adaptif bernilai cukup dan perlu ditingkatkan karena nilainya *relative* 3 (cukup); (2) Pengukuran tingkat akurasi klasifikasi menggunakan *metrics* dengan hasil *accuracy* 60% atau 0.60 di mana nilai *accuracy* ini sebagai nilai kedekatan antara nilai prediksi dengan nilai fakta, dan *error rate* 10% atau 0.10 sebagai tingkat kekeliruan pada data.

Daftar Pustaka

- [1] T. Wahyudi and P. Handayani, “Perbandingan Akurasi C4.5 dan Naive Bayes Untuk Evaluasi Kinerja Karyawan PT Catur Sentosa Adiparna,” *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022.
- [2] B. Pratiwi, A. Handayani, and Sarjana, “Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusion Matrix,” *Jurnal Informatika UPGRIS*, vol. 6, no. 2, pp. 66–75, 2020.
- [3] S. Akbar, “Analisa Faktor-Faktor yang Mempengaruhi Kinerja Karyawan,” *JIAGANIS*, vol. 3, no. 2, pp. 1–17, 2018.
- [4] G. Enzellina and D. Suhaedi, “Penggunaan Metode Principal Component Analysis dalam Menentukan Faktor Dominan,” *Jurnal Riset Matematika*, pp. 101–110, Dec. 2022, doi: 10.29313/jrm.v2i2.1192.
- [5] N. Iriadi and N. Nuraeni, “Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4.5 Untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta,” *Jurnal Teknik Komputer AMIK BSI*, vol. 2, no. 1, pp. 132–137, 2016.
- [6] Suyanto, *Data mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung, 2019.
- [7] A. Saleh, “Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,” *Citec Journal*, vol. 2, no. 3, pp. 207–217, 2015.

- [8] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *EECCIS*, vol. 7, no. 1, pp. 59–64, 2013.
- [9] J. Miharja and Suhendri, "Penerapan Data Mining Penerimaan Karyawan Menggunakan Metode *Naive Bayes Classifier*," *Proceeding SENDIU*, 2021.
- [10] I. Romli and B. Putra, "Evaluasi Penilaian Kinerja Dalam Klasifikasi Data mining Dengan Metode Naive Bayes," *Jurnal Pelita Industri*, vol. 1, no. 1, pp. 36–45, 2020.
- [11] I. Budiman, R. Ramadina, and Muliadi, "Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi," *Jurnal Jupiter*, vol. 7, no. 1, pp. 39–50, 2015.
- [12] F. Herliani and A. Kudus, "Penanganan Data Missing dengan Algoritma Multivariate Imputation By Chained Equations (MICE)," *DataMath: Journal of Statistics and Mathematics*, vol. 1, no. 1, pp. 35–42, 2023.
- [13] N. Iriadi and N. Nuraeni, "Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4.5 untuk Prediksi Kelayakan Kredit pada BANK Mayapada Jakarta," *Jurnal Teknik Komputer AMIK BSI*, vol. 2, no. 1, p. 137, 2016.