



Klasifikasi Status Kesehatan Penerimaan Calon Karyawan Menggunakan Metode *Naïve Bayes*

Rahmatina Hidayati¹, Nuryuliana², Anis Zubair³

Fakultas Teknologi Informasi, Universitas Merdeka Malang

e-mail: rahmatina.hidayati@unmer.ac.id¹, nuryuliana97@gmail.com², anis_zubair@yahoo.com³

ABSTRAK

Perusahaan memiliki kriteria tersendiri untuk memilih karyawan berkualitas, baik dari kesehatan fisik maupun mental. Tahap utama penentuan kualitas calon karyawan yaitu melakukan pemeriksaan kesehatan atau *medical check-up*. Pada tahap ini, perusahaan X masih merekap data secara manual, sehingga belum efektif dalam pengambilan keputusan. Penelitian ini menerapkan metode *Naïve Bayes* untuk mengetahui status kesehatan calon karyawan. Sampel yang digunakan sebanyak 56 data hasil *medical check-up* dengan aspek yang diperiksa antara lain tinggi badan, berat badan, tensi darah (sistolik/diastolik), mata, dan riwayat penyakit kronis. Akurasi yang diperoleh pada penelitian ini sebesar 89%.

Kata Kunci: status kesehatan karyawan, klasifikasi, *Naïve Bayes*

ABSTRACT

The company has own criteria for selecting quality employees, both from physical and mental health. The main stage of determining the quality of an applicant employee is conducting a medical check-up. At this stage, company X is still recapitulating data manually, so it is not yet effective in making decisions. This study applies the Naïve Bayes method to determine the health status of applicant employees. The samples used were 56 medical check-up results with the aspects examined, including height, weight, blood pressure (systolic / diastolic), eyes, and chronic disease. In this study produced 89 percent as the best accuracy.

Keywords: *employee health status, classification, Naïve Bayes*



PENDAHULUAN

Serangkaian tahapan harus dilalui oleh calon karyawan untuk bisa bekerja di suatu perusahaan. Tahapan tersebut antara lain tes kesehatan, tes tulis, dan tes wawancara. Kesehatan sangat dikedepankan demi menunjang kemajuan perusahaan, baik dari segi kesehatan mental maupun fisik.

Setiap perusahaan memiliki SOP (*Standard Operational Procedure*) masing-masing, akan tetapi SOP yang dijalankan antara perusahaan satu dengan yang lain tidak jauh berbeda. Untuk mengetahui tingkat kesehatan calon karyawannya, perusahaan memiliki SPK (Surat Perintah Kerja) dengan aspek yang diperiksa antara lain tinggi badan, berat badan, tensi darah (sistolik/diastolik), mata, riwayat penyakit kronis, dan lain sebagainya.

Tes kesehatan akan menghasilkan status Fit dan Tidak Fit. Hasil ini akan menjadi penentu apakah calon karyawan tersebut bisa lanjut ke tahap berikutnya. Di perusahaan X rekap data masih dilakukan secara manual, sehingga belum efektif dalam pengambilan keputusan.

Metode klasifikasi akan diterapkan untuk mengetahui status kesehatan calon karyawan. Menurut (Nisbet, Miner, & Yale, 2018) klasifikasi merupakan proses memisahkan data ke dalam beberapa kelas. Dalam klasifikasi terdapat dua fase: pelatihan dan pengujian. Fase pelatihan mencari model untuk pemisahan data, sedangkan fase pengujian untuk mengevaluasi model tersebut. Pada penelitian ini metode klasifikasi yang digunakan adalah *Naïve Bayes*.

METODE

Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa hasil *medical check-up* yang berisi data kuantitatif dan kualitatif. Keterangan beberapa atribut dan tipe datanya ditampilkan pada Tabel 1.

Tabel 1. Atribut Hasil *Medical Check-up* dan Tipe Datanya

No.	Atribut	Tipe Data
1.	Tinggi Badan	kuantitatif
2.	Berat Badan	kuantitatif
3.	Sistolik	kuantitatif
4.	Diastolik	kuantitatif
5.	Jenis Kelamin	kualitatif
6.	Mata	kualitatif
7.	Riwayat Penyakit Kronis	kualitatif
8.	Hasil Status Kesehatan	kualitatif



Hasil *medical check-up* yang terkumpul sebanyak 56. Deskripsi statistik data disajikan pada Tabel 2.

Tabel 2. Deskripsi Statistik Data

Deskripsi		Jumlah
Jenis Kelamin	L	5
	P	51
Mata	normal	55
	abnormal	1
Riwayat Penyakit Kronis	ada	11
	tidak ada	45
Status Kesehatan	fit	9
	tidak fit	47

Praproses Data

Pembersihan data merupakan prasyarat yang diperlukan demi keberhasilan mendapatkan pengetahuan dalam *database* (KDD). Tanpa pembersihan, kegunaan data pada *data mining* akan berkurang (Maimon & Rokach, 2010). Proses pembersihan data meliputi memeriksa ketidakkonsistenan data, mengisi atau membersihkan *missing values*, dan menghapus duplikat data (Han, Kamber, & Pei, 2012).

Klasifikasi

Klasifikasi merupakan proses memisahkan data ke dalam beberapa kelas. Dalam penelitian ini terdapat dua kelas, yakni Fit dan Tidak Fit. Metode klasifikasi yang diterapkan adalah *Naïve Bayes*, salah satu bentuk sederhana dari *Bayes network*.

Bayes network bisa dilihat sebagai graf asiklik terarah dengan distribusi probabilitas gabungan atas sekumpulan variabel diskrit dan stokastik. Hal ini untuk mengetahui informasi distribusi dari sampel. Sampel x akan dikelompokkan ke dalam kelas dengan probabilitas posterior tertinggi (Liao & Triantaphyllou, 2007).

Klasifikasi *Naïve Bayes* bekerja sebagai berikut (Han, Kamber, & Pei, 2012):

- (1) Misal D adalah satu set pelatihan dan label kelasnya. Setiap data ditunjukkan dengan beberapa atribut, $X = (x_1, x_2, x_3, \dots, x_n)$.
- (2) m mewakili banyaknya kelas, H_1, H_2, \dots, H_m . Jika diberikan data X , metode akan memprediksi data tersebut masuk kelas yang mana berdasarkan *posterior probability* tertinggi kondisi X . *Posterior probability* sampel x dapat dihitung dengan Persamaan (1):

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



Penjelasan mengenai peubah

$P(H | X)$: *posterior probability* H kondisi X

$P(H)$: *prior probability* dari H

$P(X | H)$: *posterior probability* X kondisi H

X : data dengan kelas yang belum diketahui

H : hipotesis data kelas spesifik

- (3) Untuk mengurangi perhitungan dalam mengevaluasi karena banyaknya atribut, maka diasumsikan nilai-nilai atribut tidak bergantung satu sama lain.

$$P(X | H_i) = P \prod_{j=1}^n (x_j | H_i) \quad (2)$$

$$= P(x_1 | H_i) * P(x_2 | H_i) * \dots * P(x_n | H_i)$$

x merupakan nilai atribut dari data. Atribut yang digunakan perlu diperhatikan apakah termasuk kualitatif atau kuantitatif. Dalam perhitungan $P(X | H_i)$ bisa mengikuti langkah berikut:

- a. Jika bertipe kualitatif maka $P(x_j | H_i)$ adalah banyaknya kondisi data yang dicari dalam kelas H_i dibagi jumlah kelas H_i . Secara umum, $P(H_i)$ dan $P(X | H_i)$ di Naïve Bayes dihitung berdasarkan frekuensi, tetapi terdapat kondisi di mana $x_j = 0$. Untuk mengatasi ini maka perlu ditambahkan *Laplace Smoothing* yang ditunjukkan pada Persamaan (3).

$$P(x_j | H_i) = \frac{1 + N(x_j, H_i)}{V + \sum_j N(x_j, H_i)} \quad (3)$$

di mana $N(x_j, H_i)$ merupakan frekuensi atribut x dari kelas H_i , dan V adalah banyaknya kelas H_i .

- b. Jika data bertipe kuantitatif, lakukan perhitungan rata-rata μ dan standar deviasi σ seperti pada Persamaan (4) dan Persamaa (5).

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4)$$

$$P(x_1 | H_i) = g(x_j, \mu_{H_i}, \sigma_{H_i}) \quad (5)$$

- (4) Perhitungan $P(H | X)$ dilakukan untuk semua kelas untuk memprediksi label kelas. Prediksi kelas terpilih merupakan hasil perhitungan maksimum.

HASIL DAN PEMBAHASAN

Pada penelitian ini sampel yang digunakan sebanyak 56 data *medical check-up* dengan atribut tinggi badan, berat badan, sistolik, diastolic, jenis kelamin, mata, dan riwayat penyakit kronis.



Keluarannya berupa status Fit dan Tidak Fit. Hasil klasifikasi dari metode Naïve Bayes ditampilkan pada Tabel 3.

Tabel 3. Hasil Klasifikasi Status Kesehatan

No	TB	BB	Sistolik	Diastolik	JK	Mata	Riwayat penyakit kronis	Hasil	Prediksi
1.	158	57	120	80	L	normal	tidak ada	Fit	Fit
2.	155	49	120	80	P	normal	tidak ada	Fit	Fit
3.	156	51	120	80	P	normal	tidak ada	Fit	Fit
4.	153	48	120	70	P	normal	tidak ada	Fit	Tidak Fit
5.	161	61	120	80	P	normal	tidak ada	Fit	Fit
6.	164	61	120	80	P	normal	tidak ada	Fit	Fit
7.	159	59	110	80	P	normal	tidak ada	Fit	Fit
8.	155	48	120	80	P	normal	tidak ada	Fit	Fit
9.	150	47	120	80	P	normal	tidak ada	Fit	Fit
10.	163	52	100	60	L	normal	ada	Tidak Fit	Tidak Fit
11.	166	50	120	80	L	normal	tidak ada	Tidak Fit	Fit
12.	165	80	120	80	L	abnormal	tidak ada	Tidak Fit	Fit
13.	153	42	90	60	P	normal	tidak ada	Tidak Fit	Tidak Fit
14.	160	68	110	80	P	normal	tidak ada	Tidak Fit	Tidak Fit
15.	158	52	90	60	P	normal	tidak ada	Tidak Fit	Tidak Fit
16.	155	38	90	60	P	normal	tidak ada	Tidak Fit	Tidak Fit
17.	154	51	100	60	P	normal	tidak ada	Tidak Fit	Tidak Fit
18.	150	45	120	80	P	normal	ada	Tidak Fit	Fit
19.	149	36	90	60	P	normal	tidak ada	Tidak Fit	Tidak Fit
20.	151	47	90	60	P	normal	ada	Tidak Fit	Tidak Fit
...
...
54.	154	47	100	60	P	normal	tidak ada	Tidak Fit	Tidak Fit
55.	164	62	100	60	P	normal	tidak ada	Tidak Fit	Tidak Fit
56.	160	52	100	70	P	normal	tidak ada	Tidak Fit	Tidak Fit

Proses klasifikasi menggunakan metode *Naïve Bayes* sebagai berikut:

1. Menghitung probabilitas masing-masing kelas. Hasil probabilitas kelas Fit dan Tidak Fit ditampilkan pada Tabel 4.

Tabel 4. Probabilitas Masing-masing Kelas

Probabilitas	
Fit	Tidak Fit
0.1607143	0.839286

$$P(\text{Fit}) = \frac{\text{banyaknya kelas Fit}}{\text{Jumlah keseluruhan kelas}} = \frac{9}{56} = 0.16$$

$$P(\text{Tidak Fit}) = \frac{\text{banyaknya kelas Tidak Fit}}{\text{Jumlah keseluruhan kelas}} = \frac{47}{56} = 0.839$$



2. Hitung nilai rata-rata dan simpangan baku atribut kuantitatif, sedangkan pada atribut kualitatif hitung *prior likelihood*. Hasil rata-rata atribut tinggi badan, berat badan, sistolik dan distolik ditampilkan pada Tabel 5.

Tabel 5. Nilai Rata-rata dan Simpangan Baku Atribut Kuantitatif

	Rata-rata		Simpangan Baku	
	Fit	Tidak Fit	Fit	Tidak Fit
TB	156.78	154.55	4.24	4.15
BB	53.44	49.30	5.96	8.88
Sistolik	118.89	100.43	3.33	9.08
Distolik	78.89	66.38	3.33	7.05

Untuk atribut jenis kelamin, mata, dan riwayat penyakit kronis, hasil perhitungan $P(x_1 | H_i)$ dengan estimasi *Laplace Smoothing* ditampilkan pada Tabel 6.

$$P(L | Fit) = \frac{1}{9} = 0.11$$

$$P(L | Tidak Fit) = \frac{4}{47} = 0.09$$

Tabel 6. *Prior Likelihood* Atribut Kualitatif

		<i>Likelihood</i>	
		Fit	Tidak Fit
JK	L	0.11	0.09
	P	0.89	0.91
Mata	Normal	1.00	0.98
	Abnormal	0.07	0.02
Riwayat Penyakit Kronis	Ada	0.07	0.23
	Tidak ada	1.00	0.77

Pada $P(Mata = Abnormal | Fit)$ dan $P(RPK = Ada | Fit)$ diterapkan *Laplace Smoothing* karena sebelumnya bernilai 0.

3. Hitung *posterior probability* $P(H | X)$ dengan menggunakan Persamaan (1). Hasil *posterior probability* ditampilkan pada Tabel 7.



Tabel 7. Hasil *Posterior Probability*

No.	<i>Posterior Probability</i>	
	Fit	Tidak Fit
1.	1.2.E-06	4.E-09
2.	8.0.E-06	9.E-08
3.	1.0.E-05	9.E-08
4.	1.5.E-07	5.E-07
5.	3.1.E-06	1.E-08
6.	1.2.E-06	3.E-09
7.	2.0.E-07	2.E-07
8.	7.0.E-06	9.E-08
9.	1.8.E-06	5.E-08
10.	4.3.E-22	1.E-08
11.	1.1.E-07	2.E-10
12.	7.7.E-13	2.E-14
13.	7.1.E-30	1.E-06
14.	1.3.E-08	3.E-08
15.	6.3.E-29	1.E-06
16.	2.1.E-30	9.E-07
17.	1.1.E-19	4.E-06
18.	8.4.E-08	1.E-08
19.	1.7.E-31	3.E-07
20.	1.1.E-30	4.E-07
...
...
54.	6.6.E-20	4.E-06
55.	1.2.E-20	1.E-07
56.	2.9.E-14	2.E-06

4. Hitung akurasi keluaran metode *Naïve Bayes*, di mana keluarannya berupa prediksi status calon karyawan yakni Fit dan Tidak Fit. Penentuan status atau kelas berdasarkan nilai *posterior probability* tertinggi.

Akurasi dihitung berdasarkan *Confusion Matrix* yang ditampilkan pada Tabel 8, di mana jumlah fakta “Fit” dengan prediksi “Fit” sebanyak 8 orang, fakta “Fit” prediksi “Tidak Fit” sebanyak 1 orang, fakta “Tidak Fit” prediksi “Fit” sebanyak 5 orang, dan fakta “Tidak Fit” prediksi “Tidak Fit” sebanyak 42 orang.

Tabel 8. Perhitungan *Confusion Matrix*

		Prediksi	
		Fit	Tidak Fit
Fakta	Fit	8	1
	Tidak Fit	5	42

Berdasarkan *Confusion Matrix* persentase akurasi dihitung dengan Persamaan (6).

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{Banyaknya prediksi benar}}{\text{total banyak prediksi}} \times 100\% \\
 &= \frac{8 + 42}{8 + 1 + 5 + 42} \times 100\% \\
 &= 89\%
 \end{aligned}
 \tag{5}$$

Dari perhitungan di atas, keberhasilan metode Naïve Bayes dalam mengklasifikasi status calon karyawan sebesar 89%.

SIMPULAN

Berdasarkan hasil penelitian ini, klasifikasi status kesehatan penerimaan calon karyawan perusahaan “X” menggunakan metode Naive Bayes menghasilkan akurasi sangat baik dengan nilai sebesar 89%. Penelitian ini bisa dikembangkan dengan sampel data yang lebih banyak dan atribut yang lebih beragam. Penelitian selanjutnya juga bisa menganalisis tingkat akurasi klasifikasi dengan menghapus salah satu atribut.

DAFTAR RUJUKAN

- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques Third Edition*. Waltham: ELSEVIER.
- Jung, Y. G., Kim, K., Lee, B., & Youn, H. (2016). Enhanced Naive Bayes Classifier for Real-time Sentiment Analysis with SparkR. *International Conference on ICT Convergence* (pp. 141-146). Seoul: IEEE.
- Liao, T. W., & Triantaphyllou, E. (2007). *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications*. World Scientific Publishing.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition*. New York: Springer.
- Nisbet, R., Miner, G., & Yale, K. (2018). *Handbook of Statistical Analysis and Data Mining Applications Second Edition*. ELSEVIER.
- Zubair, A. (2018). Penerapan Metode Naïve Bayes untuk Klasifikasi Status Gizi (Studi Kasus di Klinik Bromo Malang). *SENASIF (Seminar Nasional Sistem Informasi)* (pp. 1204-1208). Malang: Universitas Merdeka Malang.