

KLASIFIKASI DATA KINERJA KARYAWAN MENGGUNAKAN METODE HYBRID NAÏVE BAYES – K NEAREST NEIGHBOR UNTUK KENAIKAN JABATAN

Zulfiansyah^{1*}, Ade Kanianingsih¹ dan Agus Komarudin¹

¹Jurusan Informatika, Fakultas Sains Dan Informatika, Universitas Jenderal Achmad Yani
Jl. Terusan Jenderal Sudirman, Cibeber, Cimahi Selatan, Cimahi, Jawa Barat 40531

*Email: zulfiansyah17@if.unjani.ac.id

Abstrak

Ketika karyawan dipromosikan, ada kesulitan untuk mengidentifikasi siapa yang layak untuk dinaikan jabatan. Proses penilaian kinerja karyawan untuk kenaikan jabatan melalui tahap penyeleksian kriteria - kriteria seperti pendidikan, keterlibatan kerja, dan lainnya. Dari data kinerja karyawan dapat memberikan informasi untuk kenaikan jabatan pada posisi yang harus ditempati dan layak diisi. Maka dari itu, dibutuhkan suatu teknik klasifikasi data kinerja karyawan untuk kenaikan jabatan. Dalam penelitian ini akan menggunakan metode hybrid naïve bayes – k nearest neighbor. metode naïve bayes digunakan sebagai proses penyeleksian kriteria dengan mempertimbangkan probabilitas setiap kriteria. Kriteria yang telah memenuhi syarat minimum probabilitas akan dilanjutkan dengan proses klasifikasi menggunakan metode k-nearest neighbor. Penelitian ini bertujuan untuk mengetahui performa metode Hybrid Naïve Bayes – K Nearest Neighbor dalam mengklasifikasi data kinerja karyawan untuk kenaikan jabatan. Dengan mempertimbangkan penelitian terdahulu. Maka dari itu, peneliti membuat klasifikasi data kinerja karyawan menggunakan metode Hybrid Naïve Bayes – K Nearest Neighbor untuk kenaikan jabatan. Dari hasil pengujian pada Metode Hybrid Naïve Bayes – K Nearest Neighbor menunjukkan Accuracy terbaik 85%, Precision 85%, Recall 98%, F1-Score 91% dengan menggunakan $K = 9$. Oleh sebab itu, menghasilkan kriteria yaitu tingkat kerja, perfoma, keterlibatan kerja, pendidikan, tahun pelatihan, keseimbangan kehidupan kerja, dan kepuasan lingkungan untuk kenaikan jabatan.

Kata Kunci: hybrid naïve bayes k-nearest neighbor, klasifikasi, kenaikan jabatan karyawan

1. PENDAHULUAN

Kenaikan jabatan merupakan kegiatan penting dalam organisasi atau perusahaan. Karyawan akan bertanggung jawab dan berkerja dengan baik ketika ada kenaikan jabatan. Ketika karyawan dipromosikan, adanya kesulitan untuk mengidentifikasi siapa yang layak untuk dinaikan jabatan. Proses penilaian kinerja karyawan untuk kenaikan jabatan melalui tahap seleksi penilaian kriteria-kriteria seperti kompetensi, kerja tim, dan lainnya(Sunarti, 2019). Dari kriteria pendidikan, prestasi kerja, dan disiplin mempunyai pengaruh signifikan terhadap kenaikan jabatan(Hariyati, 2017). Berdasarkan dari penelitian terdahulu, kriteria – kriteria untuk kenaikan jabatan mempunyai kriteria yang jumlahnya cukup banyak. Maka dalam penelitian ini akan dilakukan penyeleksian untuk menghasilkan kriteria terbaik.

Data yang akan digunakan adalah data dari *IBM HR Analytics Employee Attrition & Performance Dataset*, yang diunduh dari *Kaggle*(Ozdemir *et al.*, 2020). Data tersebut telah digunakan oleh peneliti terdahulu untuk melakukan analisis mereka pada atrisi karyawan. Maka dalam penelitian ini data yang digunakan adalah data sekunder yang akan diolah dan diimplementasi untuk disesuaikan menjadi data primer. Dalam mengelola data tersebut, peneliti memanfaatkan klasifikasi.

Klasifikasi adalah pengelompokkan objek berdasarkan data yang dimiliki. Didalam prosesnya menjelaskan mengenai kelas data yang bertujuan untuk mendapatkan informasi yang belum diketahui(Wibawa *et al.*, 2018). Adanya klasifikasi data kinerja karyawan dapat memberikan informasi mengenai karyawan yang layak atau tidak layak dinaikan jabatan(Prabowo, 2018).

Dalam penelitian ini menggunakan metode *Hybrid Naïve Bayes – K Nearest Neighbor*. Pada metode *Naïve Bayes* akan digunakan untuk proses penyeleksian atribut yang mempertimbangkan probabilitas dari semua atribut. Atribut yang telah memenuhi ketentuan minimum probabilitas, selanjutnya dilakukan proses menggunakan metode *K Nearest Neighbor* untuk menentukan karyawan yang layak atau tidak

dinaikan jabatan(Zurroh and NIM, 2021). Pada penelitian sebelumnya, klasifikasi dengan menggunakan metode *Hybrid Naïve Bayes – K Nearest Neighbor* memperoleh akurasi yang lebih tinggi dibandingkan dengan metode *K Nearest Neighbor* yang menghasilkan akurasi 64%, sedangkan hasil klasifikasi metode *Hybrid Naïve Bayes – K Nearest Neighbor* menghasilkan akurasi 96%(Safri, Arifudin and Muslim, 2018).

Berdasarkan masalah diatas. Maka dari itu, peneliti membuat Klasifikasi Data Kinerja Karyawan Menggunakan Metode *Hybrid Naïve Bayes – K Nearest Neighbor* Untuk Kenaikan Jabatan.

2. METODOLOGI

Pada penelitian ini dilakukan beberapa tahapan dalam mencapai tujuan klasifikasi untuk kenaikan jabatan menggunakan metode *Hybrid Naïve Bayes – K Nearest Neighbor*. Berikut ini penjelasan yang dilakukan dalam mengklasifikasi data kinerja karyawan untuk kenaikan jabatan.

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data atrisi karyawan di dapatkan pada data *public* yang diperoleh pada situs <https://www.kaggle.com/> data *IBM HR Analytics Employee Attrition & Performance* dengan jumlah 1470 data dan 35 atribut.

2.2. Pra Proses

Tahap ini dilakukan seleksi data dengan memilih atribut yang berpengaruh, selanjutnya dilakukan pengelompokkan data dengan membagi data latih dan data uji, kemudian dilakukan normalisasi untuk membuat data bernilai sama dengan range 0 sampai 1 agar mempercepat proses klasifikasi, dan tahap terakhir dalam pra prose yaitu penyeleksian fitur menggunakan metode *Naive Bayes* dengan probabilitas yang ditentukan.

2.2.1 Seleksi Data

Dalam tahapan data *selection* ini dari jumlah total 35 atribut, akan dilakukan pemilihan atribut yang akan digunakan untuk menentukan faktor – faktor kenaikan jabatan karyawan. Berikut hasil pemilihan yaitu pendidikan, kepuasan lingkungan, keterlibatan kerja, tingkat kerja, kepuasan kerja, performa, kepuasan hubungan, lembur, tahun pelatihan, keseimbangan kehidupan kerja.

2.2.2 Pembagian Data Latih dan Data Uji

Pembagian data yang akan digunakan yaitu menggunakan pembagian data yang umum dipakai dalam pembagian data yaitu menggunakan perbandingan 80% : 20%(Manik, Ernawati and Nurlaili, 2021).

2.2.3 Normalisasi Data

Proses normalisasi dengan menggunakan pendekatan *Min – Max* pada semua atribut yang akan digunakan untuk mempercepat proses klasifikasi. Adapun persamaan normalisasi *Min – Max* dengan rumus pada persamaan 1 sebagai berikut(Wahanani, Prami Swari and Akbar, 2020):

$$X = \frac{X_i - Min}{Max - Min} \quad (1)$$

2.2.4 Seleksi Fitur Menggunakan Metode Naïve Bayes

Naïve Bayes merupakan suatu algoritma yang digunakan dalam proses klasifikasi dengan menggunakan metode statistika dan probabilitas (Zurroh and NIM, 2021) . Algoritma *Naïve Bayes* dilakukan dengan mencari nilai probabilitas dari setiap data atribut. Data yang memiliki probabilitas lebih besar dari nilai alpha selanjutnya akan dijumlah berdasarkan masing-masing variabel. kemudian seleksi fitur menggunakan algoritma *Naïve Bayes* dilakukan dengan cara meranking variabel yang

terpilih. Variabel yang terpilih akan diuji menggunakan algoritma *K- Nearest Neighbor* dengan menghitung $D(x,y)$ untuk setiap data yang disimpan (Safri, Arifudin and Muslim, 2018).

Pada proses klasifikasi di dalam algoritma *Naïve Bayes* dengan menggunakan data X_i kontinu, maka digunakan rumus densitas *gauss* pada persamaan 2 sebagai berikut (Yadav, Jain and Singh, 2018):

$$P(X_i) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \quad (2)$$

Persamaan 3 untuk mencari nilai rata-rata hitung (mean) sebagai berikut:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (3)$$

sedangkan Persamaan 4 untuk menghitung nilai simpangan baku (standard deviasi) sebagai berikut:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{n-1}} \quad (4)$$

2.3. Klasifikasi Menggunakan Metode K Nearest Neighbor

Dalam tahap ini setiap data akan dilakukan perhitungan menggunakan algoritma *K Nearest Neighbor*. Langkah selanjutnya yaitu dengan menghitung jarak antara data baru dan semua data yang ada di dalam data latih, sehingga akan menghasilkan nilai *K Nearest Neighbor* mayoritas (Alghifari and Wibowo, 2019; Gou *et al.*, 2019). Cara menentukan nilai K yang tepat merupakan peran yang penting dalam melakukan klasifikasi berbasis *K-Nearest Neighbor*, karena dapat mempengaruhi sensitivitas kinerja sistem itu sendiri. Secara umum prinsip kerja algoritma *K-Nearest Neighbor* adalah mencari jarak terdekat antar data dengan melakukan evaluasi terhadap tetangga terdekat dalam pelatihan(Annur, 2018). Untuk menghitung jarak dapat menggunakan rumus *Euclidean Distance* pada Persamaan 5 sebagai berikut.

$$d_i = \sqrt{\sum_{i=1}^p (X_{2i} - X_{1i})^2} \quad (5)$$

2.4. Confusion Matrix

Pengujian dalam hal ini adalah tahapan pengujian akurasi metode *Hybrid Naïve Bayes – K Nearest Neighbor* dengan *Confusion Matrix*. Pengujian akurasi metode ini dilakukan untuk mengetahui tingkat akurasi klasifikasi data kinerja karyawan yang telah dibuat dengan model pada algoritma *Hybrid Naïve Bayes – K Nearest Neighbor*, dimana sebuah matrik dari klasifikasi yang akan dibandingkan dengan kelas yang asli dari data inputan.

Ya	Tidak
True Positif (TP)	False Positif (FP)
False Negatif (FN)	True Negatif (TN)

Gambar 1 Confusion matrix

Akurasi *Confusion Matrix* merupakan rasio antara jumlah terprediksi benar dari semua data[(Hijrah, Mukhlizar and Pandria, 2020). Maka rumus untuk menghitung akurasi klasifikasi sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

$$Precision = \frac{TP}{TN + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1\ Score = 2 * \frac{(Recall * Precision)}{(Recall + Precision)} \tag{9}$$

3. HASIL DAN PEMBAHASAN

3.1 Normalisasi Data

Normalisasi data dilakukan dengan merubah rentang data menjadi lebih kecil, sehingga hasil normalisasi data didapat dari Persamaan 1 sebagai berikut:

Tabel 1 Sampel data sudah dinormalisasi

Pendidikan	Kepuasan Lingkungan	Keterlibatan Kerja	Tingkat Kerja	Kepuasan Kerja	Performa	Kepuasan Hubungan	Lembur	Tahun Pelatihan	Keseimbangan Kehidupan Kerja
0,25	0,33	0,666	0,25	1	0,66	0	1	0	0,25
0	0,66	0,33	0,25	0,33	1	1	0	0,5	0,75

3.2 Seleksi Fitur Menggunakan Metode Naïve Bayes

Sebelum dilakukan perhitungan probabilitas akan dilakukan perhitungan *Mean* dan *Standard Deviasi* dari masing-masing kategori yang ditampilkan pada Tabel 3 dan Tabel 4.

Tabel 2 Hasil perhitungan mean

Kelas	Pendidikan	Kepuasan Lingkungan	Keterlibatan Kerja	Tingkat Kerja	Kepuasan Kerja	Performa	Kepuasan Hubungan	Lembur	Tahun Pelatihan	Keseimbangan Kehidupan Kerja
Ya	0,47	0,59	0,59	0,29	0,59	0,71	0,58	0,23	0,47	0,69
Tidak	0,45	0,47	0,50	0,16	0,50	0,71	0,54	0,53	0,42	0,65

Tabel 3 Hasil perhitungan standard deviasi

Kelas	Pendidikan	Kepuasan Lingkungan	Keterlibatan Kerja	Tingkat Kerja	Kepuasan Kerja	Performa	Kepuasan Hubungan	Lembur	Tahun Pelatihan	Keseimbangan Kehidupan Kerja
Ya	0,26	0,35	0,22	0,28	0,36	0,11	0,36	0,42	0,21	0,16
Tidak	0,24	0,38	0,25	0,24	0,37	0,12	0,37	0,49	0,21	0,21

Menghitung nilai probabilitas seluruh data menggunakan rumus pada Persamaan 2 didapat nilai probabilitas seluruh data yang ditampilkan pada Tabel 6.

Tabel 4 Nilai probabilitas setiap data

Pendidikan	Kepuasan Lingkungan	Keterlibatan Kerja	Tingkat Kerja	Kepuasan Kerja	Performa	Kepuasan Hubungan	Lembur	Tahun Pelatihan	Keseimbangan Kehidupan Kerja	Kelas
0,14	0,65	0,43	0,74	0,51	0,06	0,34	0,52	0,85	0,92	1
0,14	0,16	0,79	0,43	0,51	1,05	0,34	0,52	0,85	0,92	1
0,56	0,60	0,65	0,75	0,26	1,05	0,23	0,36	0,12	0,15	0
0,56	0,25	0,62	0,64	0,59	1,05	0,55	0,36	0,81	0,77	0

Berdasarkan Tabel 6 seleksi fitur pada *Naïve Bayes* dengan memakai jumlah fitur sebanyak 7 maka dari itu dapat diketahui bahwa variabel memiliki probabilitas lebih besar dari 0.3 adalah tingkat kerja, performa, keterlibatan kerja, pendidikan, tahun pelatihan, keseimbangan kehidupan kerja, dan kepuasan lingkungan.

3.3 Klasifikasi Menggunakan Metode K Nearest Neighbor

Pada proses ini perhitungan dengan metode *K Nearest Neighbor* untuk melakukan klasifikasi terhadap objek berdasarkan dari data latih yang jaraknya paling dekat dengan data uji. Perhitungan jarak yang akan digunakan dalam penelitian ini adalah *Euclidian Distance*. Berikut adalah hasil perhitungan metode *K-Nearest Neighbor* dengan menggunakan atribut yang telah melalui tahap seleksi dapat dilihat pada Tabel 7.

Tabel 5 Hasil perhitungan menggunakan metode k nearest neighbor

Jarak	data ke-	Kelas
2	1	0
2,449489743	2	1
3	1172	3

Setelah menghitung jarak antara *Instance* maka didapatkan hasil perhitungan jarak *Euclidean Distance* sebagai tabel berikut. Hasil perhitungan ini diurutkan dari jarak yang terkecil hingga terbesar dapat dilihat pada Tabel 8.

Tabel 6 Hasil pengurutan jarak

Jarak	data ke-	Kelas
0	66	1
0	99	1
0	120	1

Berdasarkan tabel di atas dengan nilai $K=3$ diperoleh 3 jarak terdekat terdapat pada data ke 66, 99 dan 120 dan didapatkan label pada masing-masing indeks yaitu 1, 1, dan 1. Proses selanjutnya yaitu akan diambil jumlah label yang memiliki frekuensi yang mayoritas, sehingga dapat disimpulkan bahwa kelas dari data testing adalah kelas 1 atau layak dinaikan jabatan.

3.4 Hasil Pengujian

Tabel 7 Hasil prediksi confusion matrix

K	TP	FP	TN	FN
3	233	39	4	18
5	241	39	4	10
7	246	41	2	5
9	248	41	2	3

Tabel 8 Hasil Pengujian Confusion Matrix

K	Accuracy	Precision	Recall	F1-Score
3	$0,80 \times 100 = 80\%$	$0,85 \times 100 = 85\%$	$0,92 \times 100 = 92\%$	$0,89 \times 100 = 89\%$
5	$0,83 \times 100 = 83\%$	$0,86 \times 100 = 86\%$	$0,96 \times 100 = 96\%$	$0,90 \times 100 = 90\%$
7	$0,84 \times 100 = 84\%$	$0,85 \times 100 = 85\%$	$0,98 \times 100 = 98\%$	$0,91 \times 100 = 91\%$
9	$0,85 \times 100 = 85\%$	$0,85 \times 100 = 85\%$	$0,98 \times 100 = 98\%$	$0,91 \times 100 = 91\%$

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dalam mengklasifikasikan data kinerja karyawan menggunakan metode *Hybrid Naïve Bayes – K Nearest Neighbor* untuk kenaikan jabatan. Hasil pengujian *Confusion Matrix* menunjukkan *Accuracy* terbaik 85%, *Precision* 85%, *Recall* 98%, *F1-Score* 91% dengan menggunakan $k = 9$. Oleh sebab itu, menghasilkan kriteria yaitu tingkat kerja, perfoma, keterlibatan kerja, pendidikan, tahun pelatihan, keseimbangan kehidupan kerja, dan kepuasan

lingkungan untuk kenaikan jabatan karyawan dengan menggunakan metode *Hybrid Naïve Bayes – K Nearest Neighbor*.

DAFTAR PUSTAKA

- Alghifari, M.R. and Wibowo, A.P. (2019) ‘Penerapan Metode K-Nearest Neighbor Untuk Klasifikasi Kinerja Satpam Berbasis Web’, *Jurnal Teknologi dan Manajemen Informatika*, 5(1). Available at: <https://doi.org/10.26905/jtmi.v5i1.3074>.
- Annur, H. (2018) ‘Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes’, *ILKOM Jurnal Ilmiah*, 10(2), pp. 160–165. Available at: <https://doi.org/10.33096/ilkom.v10i2.303.160-165>.
- Gou, J. et al. (2019) ‘A local mean representation-based K-nearest neighbor classifier’, *ACM Transactions on Intelligent Systems and Technology*, 10(3). Available at: <https://doi.org/10.1145/3319532>.
- Hariyati, D. (2017) ‘ANALISIS FAKTOR-FAKTORYANG MEMPENGARUHI PROMOSI JABATANDINAS PERKEBUNAN PROVINSI SULAWESI SELATAN’, *Jurnal Riset Edisi XX* [Preprint].
- Hijrah, Mukhlizar, M. and Pandria, T.M.A. (2020) ‘Perbandingan Teknik Klasifikasi Untuk Memprediksi Kualitas Kinerja Karyawan’, *Jurnal Optimalisasi*, 6(1), pp. 10–21. Available at: <http://jurnal.utu.ac.id/joptimalisasi/article/view/1990>.
- Manik, G., Ernawati, I. and Nurlaili, I. (2021) ‘Analisis Sentimen Pada Review Pengguna E-Commerce Bidang Pangan Menggunakan Metode Support Vector Machine (Studi Kasus : Review Sayurbox dan Tanihub pada Google Play)’, *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, (September), pp. 64–74.
- Ozdemir, F. et al. (2020) ‘Assessing Employee Attrition Using Classifications Algorithms’, in *ACM International Conference Proceeding Series*. Association for Computing Machinery, pp. 118–122. Available at: <https://doi.org/10.1145/3404663.3404681>.
- Prabowo, (1)Nursya’bani Hendro (2018) ‘Klasifikasi Kinerja Karyawan Terhadap Penjualan Produk’, *Institut Teknologi Sepuluh Nopember* [Preprint].
- Safri, Y.F., Arifudin, R. and Muslim, A. (2018) ‘K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor’, *Scientific Journal of Informatics*, 5(1), pp. 2407–7658. Available at: <http://journal.unnes.ac.id/nju/index.php/sji>.
- Sunarti (2019) *Prediction of Employee Position Promotion Using C4.5 Algorithm (Case Study: Senayan Apartment Jakarta)*.
- Wahanani, H.E., Prami Swari, M.H. and Akbar, F.A. (2020) ‘Case based Reasoning Prediksi Waktu Studi Mahasiswa Menggunakan Metode Euclidean Distance dan Normalisasi Min-Max’, *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(6), p. 1279. Available at: <https://doi.org/10.25126/jtiik.2020763880>.
- Wibawa, A.P. et al. (2018) ‘Metode-metode Klasifikasi’, *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 3(1).
- Yaday, S., Jain, A. and Singh, D. (2018) ‘Early Prediction of Employee Attrition using Data Mining Techniques’, *Proceedings of the 8th International Advance Computing Conference, IACC 2018*, pp. 349–354. Available at: <https://doi.org/10.1109/IADCC.2018.8692137>.
- Zurroh, E.A. and NIM (2021) *KLASIFIKASI DIAGNOSA PENYAKIT DEMAM BERDARAH DENGUE MENGGUNAKAN METODE HYBRID NAÏVENAÏVE BAYES-K NEAREST NEIGHBOR*.