

Penerapan Algoritma K-Nearest Neighbor untuk Memprediksi Pengunduran Diri Karyawan

Ni Luh Komang Indira Pramesti^{a1}, Made Agung Raharja, S.Si., M.Cs.^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana

Badung, Bali, Indonesia

¹indiprames@gmail.com

²made.agung@unud.ac.id

Abstrak

Pengunduran diri karyawan adalah situasi di mana seorang karyawan meninggalkan perusahaan atas kehendak karyawan itu sendiri. Jika terlalu sering terjadi, hal ini merupakan salah satu masalah besar bagi perusahaan yang ingin terus berkembang. Oleh karena itu, penting bagi perusahaan untuk mengantisipasi hal tersebut dengan memprediksi kemungkinan karyawan akan berhenti atau tidak.

Prediksi pengunduran diri karyawan dapat dilakukan menggunakan algoritma *K-Nearest Neighbor*. Algoritma ini bekerja dengan cara menghitung jarak dari data uji ke data latih untuk menentukan tetangga terdekat. Evaluasi akhir dilakukan menggunakan *confusion matrix*. Hasil klasifikasi menggunakan *K-Nearest Neighbor* dengan nilai $K = 5$ menghasilkan nilai *accuracy* sebesar 85%, *precision* sebesar 50%, dan *recall* sebesar 33,33%.

Kata Kunci: *K-Nearest Neighbor, Pengunduran Diri, Prediksi, Klasifikasi, Karyawan*

1. Pendahuluan

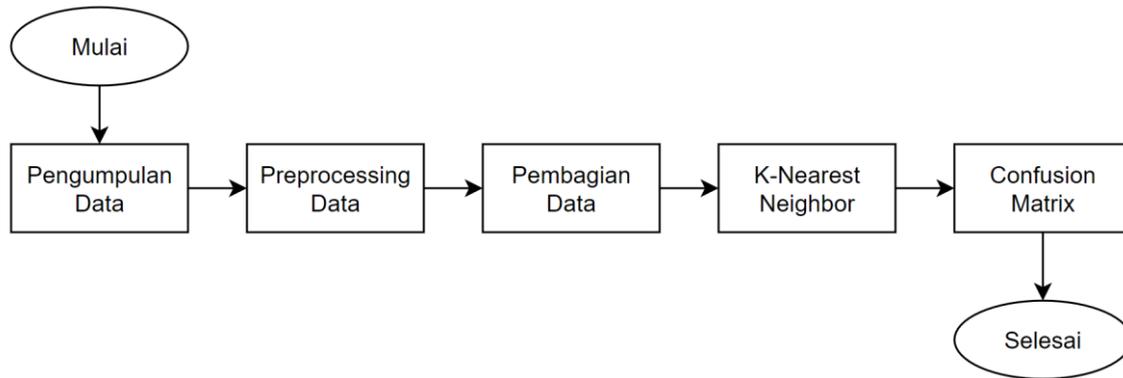
Pengunduran diri karyawan adalah situasi di mana seorang karyawan meninggalkan perusahaan atas kehendak karyawan itu sendiri. Jika terlalu sering terjadi, hal ini merupakan salah satu masalah besar bagi perusahaan yang ingin terus berkembang karena biaya dan waktu yang dibutuhkan untuk mencari kandidat baru yang sesuai dengan standar perusahaan tidaklah sedikit [1]. Dalam beberapa kasus, terdapat karyawan yang mengundurkan diri tanpa memberitahu perusahaan terlebih dahulu. Hal ini dapat berujung pada kerugian bagi perusahaan dalam bentuk penurunan produktivitas dan efisiensi kerja [2]. Oleh karena itu, penting bagi perusahaan untuk mengantisipasi hal tersebut dengan memprediksi kemungkinan karyawan akan berhenti atau tidak. Dengan mengetahui kemungkinan tersebut, perusahaan dapat mulai mencari kandidat baru lebih awal tanpa khawatir terhadap penurunan produktivitas.

Prediksi tersebut dapat dilakukan dengan menerapkan algoritma *machine learning*, salah satunya adalah *K-Nearest Neighbor*. *K-Nearest Neighbor* merupakan salah satu algoritma yang banyak digunakan dalam kasus memprediksi sesuatu. Algoritma ini termasuk ke dalam algoritma *supervised learning*. Prinsip kerjanya adalah data yang diuji akan termasuk ke dalam label terbanyak yang terdekat dari data uji [3].

Penelitian serupa sebelumnya pernah dilakukan dengan judul Prediksi Pengunduran Diri Karyawan Perusahaan "Y" Menggunakan *Random Forest* [4]. Penelitian tersebut menghasilkan nilai akurasi sebesar 87% dan *error* sebesar 13%. Dalam penelitian ini, akan dilakukan prediksi pengunduran diri karyawan menggunakan algoritma *K-Nearest Neighbor*.

2. Metode Penelitian

Proses penelitian dimulai dari pengumpulan data, *preprocessing* data, pembagian data, implementasi algoritma *K-Nearest Neighbor*, dan diakhiri dengan evaluasi menggunakan *confusion matrix*.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Data pengunduran diri karyawan yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari situs Kaggle. Data ini terdiri dari 8 atribut (Gender, Satisfaction, Business Travel, Department, EducationField, Salary, Home-Office, Attrition) dari 100 karyawan di mana sebanyak 18 karyawan meninggalkan perusahaan dan 82 karyawan memutuskan untuk tetap berada di perusahaan. Deskripsi dari masing-masing atribut dapat dilihat di tabel 1.

Tabel 1. Atribut Data

Atribut	Deskripsi	Nilai kategori
Gender	Gender pegawai	- Male - Female
Satisfaction	Tingkat kepuasan pegawai terhadap lingkungan kerja dan perusahaan	- High - Medium - Low
Business Travel	Seberapa sering karyawan melakukan perjalanan bisnis	- Rare - Frequent - No
Department	Departemen pegawai	- R&D - Sales
EducationField	Kualifikasi bidang pendidikan	- Engineering - Medical - Marketing - Other - Technical Degree
Salary	Golongan gaji karyawan	- High - Medium - Low
Home-Office	Jarak dari rumah ke kantor	- Near - Far
Attrition	Pegawai meninggalkan perusahaan atau tidak	- Yes - No

2.2 Preprocessing Data

Sebelum dapat digunakan, data sebaiknya dibersihkan terlebih dahulu. Pembersihan data pengunduran diri karyawan meliputi menghilangkan data yang memiliki nilai kosong dan mengubah data kategoris menjadi numerik agar lebih mudah diinterpretasikan oleh model *machine learning*.

2.3 Pembagian Data

Setelah data dibersihkan, langkah selanjutnya yang dilakukan adalah memisahkan data menjadi 2, yaitu data latih dan data uji. Data latih terdapat sebanyak 80% dari seluruh data dan data uji merupakan 20% sisanya.

2.4 K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* (KNN) merupakan algoritma klasifikasi yang termasuk ke dalam *supervised learning*. Cara kerjanya adalah dengan menghitung jarak dari data uji ke data latih untuk menentukan tetangga terdekat. Setelah itu, kategori mayoritas tetangga terdekat merupakan prediksi dari data uji tersebut. Jarak antar tetangga dapat dihitung menggunakan rumus *Euclidean* [5].

$$D(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Keterangan:

D = jarak antara titik

p = data latih

q = data uji

i = nilai atribut

n = dimensi atribut

Selengkapnya, berikut langkah-langkah menggunakan algoritma KNN [5]:

- Menentukan parameter K (jumlah tetangga paling dekat).
- Menghitung jarak antara data uji dan data latih menggunakan rumus *Euclidean*.
- Kemudian mengurutkan jarak yang terbentuk.
- Menentukan jarak terdekat sampai urutan K.
- Memasangkan kelas yang bersesuaian.
- Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

2.5 Confusion Matrix

Confusion matrix merupakan metode evaluasi yang banyak digunakan untuk mengukur kinerja suatu algoritma klasifikasi. *Confusion matrix* mengandung informasi perbandingan hasil klasifikasi yang diprediksi sistem dan hasil yang seharusnya. Pada pengukuran ini, terdapat empat komponen yang dimiliki, yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) [6]. Seperti yang dapat dilihat pada gambar 2 [7], kolom menunjukkan hasil klasifikasi yang sebenarnya dan baris menunjukkan hasil yang diprediksi oleh sistem.

		ACTUAL VALUES	
		POSITIVE	NEGATIVE
PREDICTED VALUES	POSITIVE	TP	FP
	NEGATIVE	FN	TN

Gambar 2. Confusion Matrix

Dari *confusion matrix* tersebut, dapat dihitung nilai *accuracy*, *precision*, dan *recall* menggunakan rumus berikut [8]:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$precision = \frac{TP}{TP+FP} \quad (3)$$

$$recall = \frac{TP}{TP+FN} \quad (4)$$

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini adalah data pengunduran diri karyawan yang terdiri dari 100 baris data dan 8 atribut. Karena tidak ada data dengan nilai yang kosong, maka semua data tersebut bisa digunakan. Kemudian, data yang aslinya memiliki nilai kategoris pada semua kolom diubah menjadi nilai integer agar lebih mudah diproses. Perbedaan data sebelum dan sesudah melalui *preprocessing* dapat dilihat pada tabel 2 dan 3.

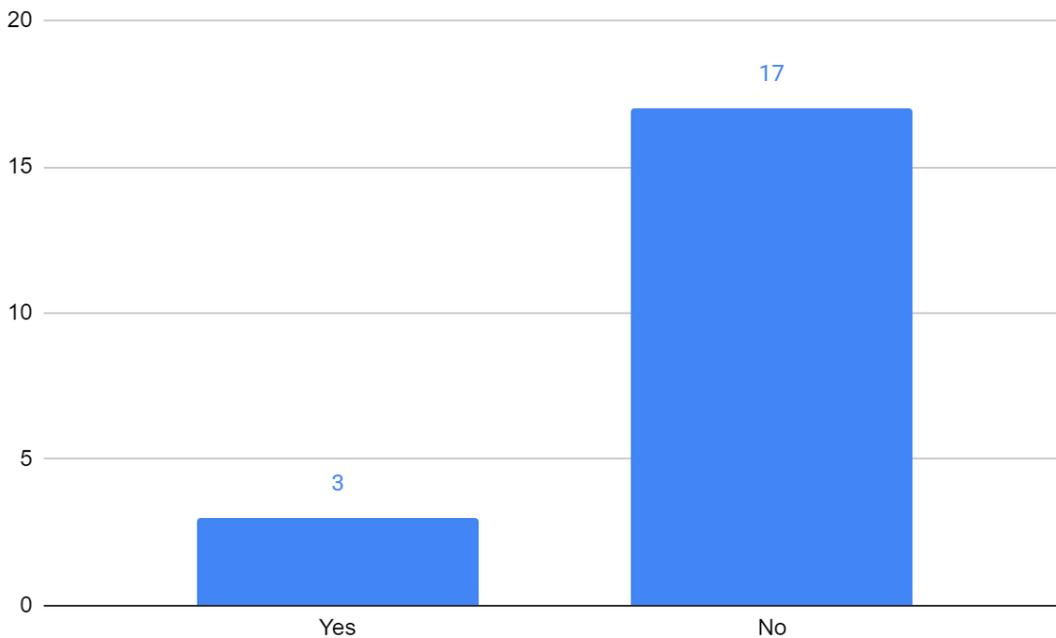
Tabel 2. Data Sebelum *Preprocessing*

Gender	Satisfaction	Business Travel	Department	EducationField	Salary	Home - Office	Attrition
Female	High	Rare	Sales	Engineering	Medium	Near	Yes
Male	Low	Frequent	Sales	Engineering	Low	Near	Yes
Male	Medium	Rare	R&D	Other	Medium	Far	No
Female	Low	Frequent	R&D	Engineering	Medium	Far	Yes
Male	Medium	Rare	R&D	Medical	Medium	Far	No

Tabel 3. Data Setelah *Preprocessing*

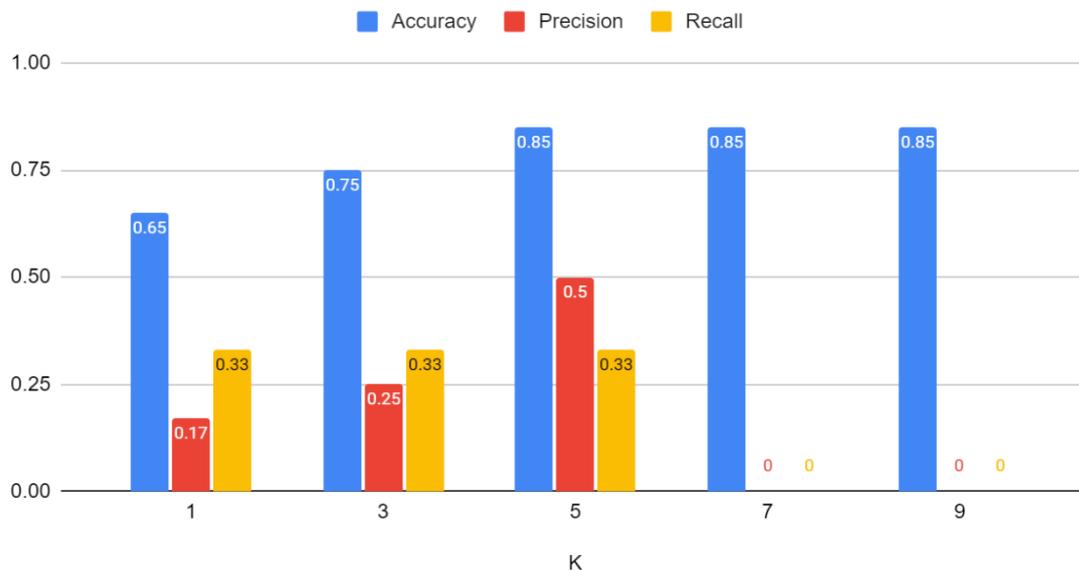
Gender	Satisfaction	Business Travel	Department	EducationField	Salary	Home-Office	Attrition
0	3	1	1	0	2	1	1
1	1	2	1	0	1	1	1
1	2	1	0	3	2	2	0
0	1	2	0	0	2	2	1
1	2	1	0	1	2	2	0

Sebelum mengimplementasikan algoritma KNN, data dibagi menjadi dua, yaitu data latih sebanyak 80% dan data uji sebanyak 20%. Data uji terdiri dari 3 data karyawan yang meninggalkan perusahaan dan 17 data karyawan yang tidak meninggalkan perusahaan seperti yang ditunjukkan pada gambar 3.



Gambar 3. Perbandingan Pengunduran Diri Karyawan pada Data Uji

Pemodelan *K-Nearest Neighbor* dilakukan sebanyak 5 kali dengan nilai K yang berbeda. Nilai K ditentukan dengan menguji bilangan ganjil dari 1 sampai 9 kemudian membandingkan hasil metriknya yaitu *accuracy*, *precision*, dan *recall*. Bilangan genap tidak digunakan untuk menghindari hasil jumlah tetangga yang seri. Perbandingan hasil metrik untuk masing-masing nilai K dapat dilihat pada gambar 4.



Gambar 4. Perbandingan Metrik Nilai K

Berdasarkan grafik tersebut, pemodelan terbaik diperoleh dengan menggunakan K = 5 karena memiliki nilai *accuracy*, *precision*, dan *recall* tertinggi dibandingkan nilai K lainnya. Dari 20 data yang diuji, diperoleh *confusion matrix* yang terdiri dari nilai *True Positive* (TP) sebanyak 1, *False Positive* (FP) sebanyak 1, *False Negative* (FN) sebanyak 2, dan *True Negative* (TN) sebanyak 16. Berdasarkan *confusion matrix* tersebut, diperoleh hasil *accuracy* sebesar 85%, *precision* sebesar 50%, dan *recall* sebesar 33,33%.

4. Kesimpulan

Dalam penelitian ini, prediksi pengunduran diri karyawan dilakukan menggunakan algoritma *K-Nearest Neighbor* dengan nilai K = 5. Kinerja algoritma dinilai menggunakan *confusion matrix* untuk menghitung nilai *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Dari *confusion matrix* tersebut dapat dihitung nilai *accuracy*, *precision*, dan *recall*. Diperoleh hasil akhir yaitu nilai *accuracy* sebesar 85%, *precision* sebesar 50%, dan *recall* sebesar 33,33%. Nilai *accuracy* yang lumayan tinggi tersebut menandakan bahwa model yang telah dibuat sudah cukup baik dalam memprediksi data secara keseluruhan. Namun, nilai *precision* dan *recall* yang rendah menunjukkan bahwa model tidak bisa memprediksi kelas positif, dalam hal ini karyawan yang mengundurkan diri, dengan baik.

Diharapkan penelitian selanjutnya dapat mengembangkan algoritma yang lebih baik untuk memprediksi pengunduran diri karyawan agar dapat memprediksinya dengan lebih akurat dan tepat.

Daftar Pustaka

- [1] N. N. Rahma, "Pengunduran Diri Karyawan Massal Pasca-Lebaran Diperkirakan Meningkat! Ini Saran EngageRocket," 2022. <https://wartaekonomi.co.id/read407407/pengunduran-diri-karyawan-massal-pasca-lebaran-diperkirakan-meningkat-ini-saran-engagerocket> (accessed Oct. 02, 2022).

- [2] E. I. Larasati, "KERUGIAN PERUSAHAAN AKIBAT PENGUNDURAN DIRI PEKERJA WAKTU TERTENTU TANPA ADANYA PEMBERITAHUAN KEPADA PERUSAHAAN," *Jurist-Diction*, vol. 2, no. 1, pp. 112–130, 2019.
- [3] N. K. S. P. Rahayu and I. K. A. Mogi, "Implementation of K-Nearest Neighbor Algorithm in Heart Disease Classification," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 10, no. 1, pp. 39–44, 2021.
- [4] D. D. E. Manurung, F. Sandi, F. Akbardipura, H. Ashfahan, and D. S. Prasvita, "Prediksi Pengunduran Diri Karyawan Perusahaan 'Y' Menggunakan Random Forest," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, Sep. 2021, pp. 202–213.
- [5] D. Cahyantia, A. Rahmayania, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 39–43, Jul. 2020.
- [6] Karsito and S. Susanti, "PENGAJUAN KREDIT RUMAH DENGAN ALGORITMA NAÏVE BAYES DI PERUMAHAN AZZURA RESIDENCIA," *SIGMA – Jurnal Teknologi Pelita Bangsa*, vol. 9, no. 3, pp. 43–48, 2019.
- [7] A. Bhandari, "Everything you Should Know about Confusion Matrix for Machine Learning," 2020. <https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/> (accessed Oct. 03, 2022).
- [8] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, Sep. 2021.