Vol 3.No.1 2024 Hal 277-288

Model Klasifikasi Kesehatan Karyawan Berbasis *Naïve Bayes Classifier* dan *K-Nearest Neighbor*

Zefri Dafianto[1], Yogi Yunefri[2], Digdaya Arief [3]

[1,2,3]Program Studi Teknik Informatika Fakultas Ilmu Komputer
Universitas Lancang Kuning
[1,2,3]Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015
e-mail: [1]zefri@unilak.ac.id, [2]yogiyunefri@unilak.ac.id, [3]digdayaarief@gmail.com

Abstrak

Penelitian ini memfokuskan pada pentingnya kesejahteraan karyawan sebagai pilar fundamental dalam keberhasilan perusahaan serta bagaimana PT. Serba Dinamik Indonesia telah mengenali dan berkomitmen untuk meningkatkan kualitas hidup para karyawannya dengan menjaga kesehatan mereka. Menyadari era digital yang memudahkan pengumpulan data kesehatan, perusahaan mengerti bahwa evaluasi manual terhadap informasi tersebut tidaklah efisien. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem yang dapat secara otomatis mengklasifikasikan kondisi kesehatan karyawan PT. Serba Dinamik Indonesia. Dengan menerapkan algoritma Naïve Bayes Classifier dan K-Nearest Neighbor, menggunakan Google Colaboratory, penelitian ini berhasil mengolah data absensi sebanyak 9910 karyawan untuk pelatihan data dan 10 data uji. Hasil klasifikasi menunjukkan bahwa metode K-Nearest Neighbor berhasil dengan tingkat akurasi 80%, sementara metode Naïve Bayes Classifier mencapai 70%. Implementasi ini mengindikasikan bahwa pendekatan K-Nearest Neighbor mungkin lebih efektif dalam mengklasifikasikan kondisi kesehatan karyawan secara otomatis bagi PT. Serba Dinamik Indonesia, yang dapat menjadi langkah penting dalam menjaga kesejahteraan mereka dan pada gilirannya, meningkatkan produktivitas serta suasana kerja yang harmonis di perusahaan tersebut.

Kata Kunci*: Data Mining, Klasifikasi, Naïve Bayes Classifier, K-Nearest Neighbor, Google Colaboratory*

Abstract

This research focuses on the importance of employee welfare as a fundamental pillar in a company's success and how PT. Serba Dinamik Indonesia has recognized and committed to enhancing the quality of life for its employees by prioritizing their health. Understanding the digital era's convenience in collecting health data, the company acknowledges that manual evaluation of this information is inefficient. Therefore, this study aims to develop a system capable of automatically classifying the health conditions of PT. Serba Dinamik Indonesia's employees. By applying Naïve Bayes Classifier and K-Nearest Neighbor algorithms using Google Colaboratory, the research successfully processed attendance data from 9910 employees for training and used 10 test data. The classification results show that the K-Nearest Neighbor method achieved an accuracy rate of 80%, while the Naïve Bayes Classifier reached 70%. This implementation suggests that the K-Nearest Neighbor approach might be more effective in automatically classifying employee health conditions for PT. Serba Dinamik Indonesia, which could be a crucial step in safeguarding their welfare and, consequently, enhancing productivity and fostering a harmonious work environment within the company.

Keywords*: Data Mining, Classification, Naïve Bayes Classifier, K-Nearest Neighbor, Google Colaboratory*

1. PENDAHULUAN

PT. Serba Dinamik Indonesia mengutamakan kesejahteraan karyawan demi kesuksesan perusahaan. Dalam memperkuat program kesehatan, mereka mengenali perlunya pendekatan yang lebih tepat dalam mengidentifikasi kondisi kesehatan karyawan. Dalam era digital, data kesehatan karyawan dapat dikumpulkan dari berbagai sumber, namun, evaluasi manual kurang efisien. Penelitian ini bertujuan mengembangkan sistem otomatis untuk mengklasifikasikan kondisi kesehatan karyawan menggunakan algoritma Naïve Bayes Classifier dan K-Nearest Neighbor (KNN) dengan Google Colaboratory. *Naïve Bayes Classifier* atau sering disebut *Bayesian Classification* adalah metode pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Naïve Bayes Classifier* (NBC) terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar [1]. Sementara itu, metode *K-Nearest Neighbors* (KNN) merupakan sebuah metode untuk melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. *K-Nearest Neighbors* termasuk algoritma *supervised learning*, Prinsip kerja *K-Nearest Neighbors* (KNN) sendiri adalah mencari jarak terdekat antara data yang dievaluasi dengan (K) tetangga terdekatnya dalam data pelatihan (Yuliansyah et al., 2022).

Analisis ini diharapkan membantu PT. Serba Dinamik Indonesia memantau kesehatan karyawan secara efisien, mendeteksi dini masalah kesehatan, serta memberikan perawatan yang tepat waktu. Ini dapat meningkatkan kualitas hidup karyawan, mengurangi absensi kerja akibat kesehatan, dan pada akhirnya, meningkatkan produktivitas mereka. Penelitian ini juga memberikan wawasan tentang penerapan algoritma machine learning dalam konteks perusahaan, berguna untuk manajemen SDM dan perencanaan strategis. Hasilnya dapat berkontribusi pada pengembangan teknologi berbasis data dalam perawatan kesehatan untuk perusahaan lain yang memiliki fokus serupa pada kesejahteraan karyawan.

Dalam beberapa penelitian terkait penggunaan metode klasifikasi seperti K-Nearest Neighbors (KNN) dan Naïve Bayes Classifier (NBC), berbagai perbandingan dan implementasi dilakukan pada beragam bidang. Misalnya, penelitian [2] yang membandingkan performa KNN dan NBC dalam klasifikasi status gizi balita menunjukkan keunggulan KNN dengan selisih f1 score yang cukup besar, sedangkan penelitian sebelumnya seperti yang dilakukan [3] menggunakan NBC untuk analisis survei kesehatan mental dengan akurasi sebesar 72%. Sementara pada klasifikasi penyakit jantung, penelitian [4] menyimpulkan NBC memiliki akurasi lebih tinggi daripada KNN. Begitu juga pada penelitian lainnya, seperti dalam pengukuran kinerja sales marketing [5] yang menunjukkan nilai akurasi 91,67% dari algoritma NBC yang digunakan. Meskipun demikian, hasil dari penelitian [6] menunjukkan bahwa C4.5 berhasil mereduksi jumlah atribut dan memiliki akurasi tertinggi dalam klasifikasi penerima bantuan Program Keluarga Harapan. Sementara itu, dalam analisis penyakit stroke, penelitian [7] menegaskan bahwa KNN memberikan hasil akurasi yang lebih tinggi dibandingkan NBC.

Berbagai penelitian lainnya, seperti klasifikasi jamur beracun [8] menunjukkan bahwa KNN memiliki akurasi yang lebih baik daripada NBC, namun kedua algoritma tersebut masih memberikan performa yang sangat baik dengan nilai di atas 90%. Dalam konteks meramalkan status pekerjaan alumni, penelitian [9] menunjukkan bahwa Naïve Bayes memiliki akurasi yang sedikit lebih baik dibandingkan KNN. Dalam berbagai penelitian terkait penggunaan metode klasifikasi seperti K-Nearest Neighbors (KNN) dan Naïve Bayes Classifier (NBC), perbandingan serta implementasi dilakukan pada beragam bidang. Misalnya, penelitian [10] menggunakan Naïve Bayes untuk pemeriksaan kesehatan

harian di tempat kerja guna menjaga keselamatan pekerja. Sementara penelitian [1] fokus pada prediksi keberhasilan immunotherapy dalam menyembuhkan penyakit kanker kulit dengan menemukan bahwa model KNN memiliki akurasi lebih tinggi daripada Naïve Bayes. Dari sekian banyak penelitian tersebut, terlihat variasi hasil tergantung pada konteks dan dataset yang digunakan.

Klasifikasi adalah proses identifikasi pola atau fungsi yang memisahkan kelas data berbeda untuk memprediksi kelas suatu objek tanpa label. Proses ini melibatkan pembentukan model yang mengelompokkan data ke dalam kategori berbeda berdasarkan aturan atau fungsi khusus, seperti aturan "jika-maka", pohon keputusan, atau formula matematis [5]. Untuk melakukan klasifikasi, diperlukan data seperti usia, kadar gula darah, kolesterol, yang diolah menggunakan metode klasifikasi tertentu untuk mendapatkan prediksi yang diinginkan [4].

Naïve Bayes yakni pengategorian dengan prosedur probalitas dan statistik yang dijumpai oleh akademikus kebangsaan Inggris Thomas Bayes, yang memproyeksikan probabilitas di waktu depan bersumber pada pengalaman yang telah berlangsung (Akmal et al., 2023). Naïve Bayes Classifier didasarkan kepada teorema Bayes yang memiliki bentuk umum sebagai berikut (Cahyanti et al., 2021):

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

K-Nearest Neighbor merupakan metode klasifikasi yang menggunakan ketetanggaan sebagai nilai prediksi terhadap nilai objek (Maricar & Pramana, 2019). Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada (Cahyanti et al., 2021). Adapun rumus dari metode K-Nearest Neighbors adalah sebagai berikut (Yuliansyah et al., 2022):

$$d(x,y) = \sqrt{\sum_{i=1}^{r}(x_{ik} - y_{ik})^2}$$

Confusion Matrix merupakan sebuah matriks yang menampilkan visualisasi kinerja dari algoritma klasifikasi menggunakan data dalam matriks yang membagi klasifikasi prediksi dalam bentuk *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Tabel berikut menjelaskan perbandingan dari *Confusion Matrix* (Batubara et al., 2023):

Tabel 1. Confusion Matrix

Nilai Prediksi	**Nilai Aktual**	
	+	**-**
+	True Positive (TP)	False Positive (FP)
-	False Negatives(FN)	True Negative (TN)

Untuk menghitung digunakan persamaan di bawah ini (Dina et al., 2023):

Sensitivity= TP/P Precision = TP/(TP+FP)
Specificity = TN/P Accuracy = Sensitivity P/(P+N) + Specificity N/(P+N)

Keamanan, Kesehatan, dan Keselamatan Kerja (K3) merupakan salah satu hal penting yang wajib diterapkan oleh semua perusahaan. Hal ini juga tertuang dalam Undang-Undang Ketenagakerjaan No. 13 Tahun 2003 pasal 87. Menurut UU No. 1 Tahun 1970 Tentang

Keselamatan Kerja, tujuan dari K3 adalah mencegah terjadinya kecelakaan dan sakit dikarenakan pekerjaan. Pengertian K3 adalah bidang yang berhubungan dengan keselamatan, kesehatan, dan kesejahteraan manusia yang bekerja pada sebuah institusi ataupun lokasi proyek. Arti K3 (Keamanan, Kesehatan, dan Keselamatan Kerja) secara khusus dapat dibagi menjadi dua, yaitu (Wahyuningsih et al., 2021):

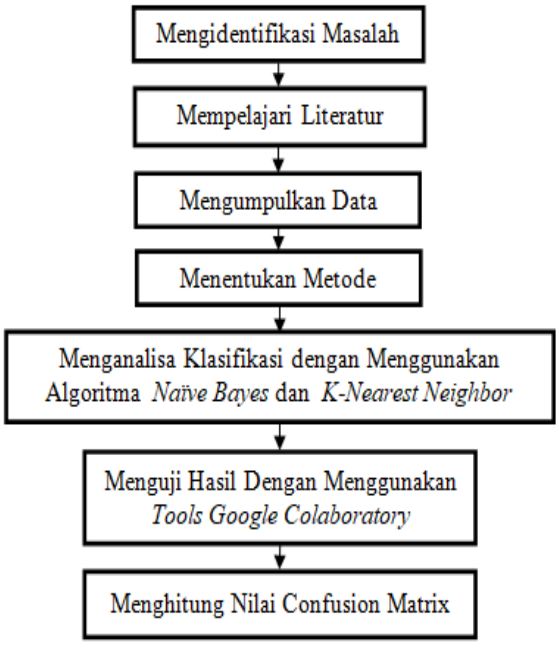
a. Pengertian K3 secara keilmuan; K3 merupakan ilmu pengetahuan dan penerapannya dalam upaya mencegah terjadinya kecelakaan dan penyakit akibat kerja.
b. Pengertian K3 secara filosofis; suatu upaya yang dilakukan untuk memastikan keutuhan dan kesempurnaan jasmani dan rohani tenaga kerja pada khususnya, dan masyarakat pada umumnya terhadap hasil karya dan budaya menuju masyarkat adil dan makmur.

Kesehatan adalah kebutuhan dasar yang tak terkecuali bagi siapa pun, mengabaikan jenis kelamin, usia, suku, atau status sosial [12]. WHO memandang kesehatan sebagai sumber daya hidup sehari-hari, sementara UU Kesehatan mengartikulasikan kesehatan sebagai kesejahteraan fisik, mental, dan sosial, bukan hanya sebagai ketiadaan penyakit, cacat, atau kelemahan [13].

Python merupakan bahasa pemrograman serbaguna yang dikembangkan oleh Guido van Rossum pada tahun 1990. Python merupakan kelanjutan dari bahasa pemrograman ABC. Nama python dipilih oleh Guido sebagai nama bahasa ciptaannya karena kecintaan Guido pada acara televise Monty Python's Flying Circus (Yuliansyah et al., 2022). Python merupakan bahasa pemrograman yang freeware atau perangkat bebas dalam arti sebenarnya, tidak ada batasan dalam penyalinannya atau mendistribusikannya. Lengkap dengan source codenya, debugger dan profiler, antarmuka yang terkandung di dalamnya untuk pelayanan antarmuka, fungsi sistem, GUI (antarmuka pengguna grafis), dan basis datanya (Rahma et al., 2021).

2. METODE PENELITIAN

Adapun beberapa tahapan pada penelirian ini adalah sebagai berikut :



Gambar 1. Tahapan Penelitian

Berdasarkan gambar 1 diatas, langkah-langkah penelitian ini diuraikan sebagai berikut:

a. Identifikasi Masalah: Fokus pada klasifikasi kondisi kesehatan karyawan PT. Serba Dinamik Indonesia dengan pendekatan datamining.
b. Review Literatur: Studi literatur dari jurnal ilmiah, modul pembelajaran, dan buku Data Mining, terutama yang berhubungan dengan Algoritma Naïve Bayes dan K-Nearest Neighbor, sebagai panduan dalam penelitian.
c. Kumpulkan Data: Pengumpulan data Absensi dan *Daily Check Up* karyawan divisi HES PT. Serba Dinamik Indonesia, serta informasi perusahaan seperti sejarah, visi, dan struktur organisasi.
d. Penentuan Metode: Memilih metode klasifikasi data, yang dalam hal ini adalah Naïve Bayes dan K-Nearest Neighbor.
e. Analisis dengan Algoritma: Pengolahan dan desain pola data yang disesuaikan untuk penggunaan metode yang telah dipilih, diharapkan menghasilkan hasil sesuai ekspektasi.
f. Pengujian dengan Google Colaboratory: Analisis dan implementasi data dengan aturan Algoritma Naïve Bayes dan K-Nearest Neighbor, menggunakan Microsoft Office Excel dan Google Colaboratory untuk mendapatkan pengetahuan yang diharapkan.
g. Perhitungan Confusion Matrix: Langkah akhir, menghitung akurasi prediksi kedua metode dengan Confusion Matrix di Google Colaboratory.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data

Analisis data pada penelitian ini menggunakan dua metode yaitu *Naïve Bayes dan K-Nearest Neighbor* yang nantinya hasil analisis akan dibandingkan dengan perhitungan dan pengujian dengan *Tools* Google Colaboratory untuk melihat metode mana yang paling optimal dalam proses pengklasifikasian dalam melihat klasifikasi kondisi kesehatan karyawan pada PT. Serba Dinamik indonesia. Adapun data yang didapat adalah sebagai berikut (Selengkapnya pada Lampiran 1):

TABEL 2. Data Absensi dan Pengecekkan Kesehatan Karyawan PT.SDI

No.	Tanggal	Nama	Kondisi Kesehatan Lain			Pemeriksaan Kesehatan Harian						
			Keterangan	Keluhan	Obat Yang Dikonsumsi	Suhu Tubuh	Frekuensi Nadi	Tekanan Sistolik	Tekanan Diastolik	Frekuensi Nafas	Gula Darah	Romberg Test
1	01/01/2023	WAHYU EKO	SAKIT	Y	0	37,70	65	137	83	12	197,94	0
2	01/01/2023	ALMIZAN	SAKIT	Y	0	38,90	104	100	75	26	117,63	1
3	01/01/2023	RADINO	SAKIT	Y	1	39,40	121	98	76	11	85,31	0
4	01/01/2023	SATRIA ADE	SAKIT	Y	1	37,70	130	146	94	16	228,01	0
5	01/01/2023	ASRUL	SAKIT	Y	0	39,50	129	105	64	26	103,49	1
..	..	..	..	..	..	..	..	..	..	..	..	..
9906	20/09/2023	ZAKI FAJRUL	KERJA	N	0	35,60	121	113	74	24	219,38	0
9907	20/09/2023	ROMES YUNES	KERJA	N	0	36,60	106	118	78	20	173,3	0
9908	20/09/2023	ZULMUDASIR	KERJA	N	0	37,10	72	105	61	24	72,72	0
9909	20/09/2023	MUHAMMAD IHSAN	KERJA	N	0	36,50	107	129	88	29	77,81	0
9910	20/09/2023	APRIADI	KERJA	N	1	35,90	88	144	98	21	123,45	0

Data pada tabel 1 diatas merupakan data absensi dan pengecekkan kondisi kesehatan karyawan pada PT. Serba Dinamik Indonesia yang di rekap dari tanggal 01 Januari hingga 20 September 2023. Data berisikan tanggal, nama karyawan, keterangan absensi, serta pengecekkan kesehatan.

3.2 Praproses Data

Pada praproses data ini dilakukan dengan mengambil data mentah dari tabel 5.1 lalu direkap dalam bentuk data yang tersimpan dengan format *Excel* (.xls), yang nantinya akan dijadikan sebuah data baru (*data training*) yang terdiri dari Data Kode Absensi Karyawan, Keluhan, Obat Yang Dikonsumsi, Suhu Tubuh, Frekuensi Nadi, Tekanan Sistolik, Tekanan Diastolik, Frekuensi Nafas, Gula Darah, Romberg Test, dan Keterangan Absensi (Label) sebagai bahan perhitungan. Adapun data yang telah melawati proses transformasi dari data mentah menjadi *data training* yang siap untuk digunakan dalam proses perhitungan pada kedua metode adalah sebagai berikut (Selengkapnya pada Lampiran 2) :

TABEL 3. *Data Training*

Kode	Keluhan	Obat Yang Dikonsumsi	Suhu Tubuh	Frekuensi Nadi	Tekanan Sistolik	Tekanan Diastolik	Frekuensi Nafas	Gula Darah	Romberg Test	Keterangan
A00001	1	0	1	0	1	1	0	1	0	SAKIT
A00002	1	0	1	1	0	0	3	0	1	SAKIT
A00003	1	1	1	1	0	0	1	0	0	SAKIT
A00004	1	1	1	1	2	2	0	2	0	SAKIT
A00005	1	0	1	1	0	0	3	0	1	SAKIT
..	..	..	..	..	..	..	..	..	..	..
A09906	0	0	0	1	0	0	2	2	0	KERJA
A09907	0	0	0	1	0	0	0	0	0	KERJA
A09908	0	0	0	0	0	0	2	0	0	KERJA
A09909	0	0	0	1	1	1	3	0	0	KERJA
A09910	0	1	0	0	2	2	2	0	0	KERJA

Pada Tabel 3 terdapat 9910 data absensi yang telah melalui praproses data yang dijadikan sebagai *data training*. Posisi Keterangan Absensi (Label) dipindahkan menjadi urutan data paling belakang. Ada pun keterangan dari tabel 3 dapat dilihat pada tabel 4 berikut :

TABEL 4. Kode dan Keterangan Data

Kode	Jenis Data	Keterangan (Kategori)
Kode Absensi	Kode	Diawali dengan huruf A untuk menyatakan Absensi
Keluhan	Variabel	Tidak Ada (0), Ada (1)
Obat Yang Dikonsumsi	Variabel	Tidak Ada (0), Ada (1)
Suhu Tubuh	Variabel	Normal jika <=37,2 (0), Tidak Normal jika >37,2 (1)
Frekuensi	Variabel	Normal jika <=100 (0), Tidak Normal jika >100 (1)
Tekanan Sistolik	Variabel	Normal jika <=120 (0), Prehipertensi jika <140 (1), Hipertensi Stadium I jika <160 (2), Hipertensi Stadium II jika >=160 (3)
Tekanan Diastolik	Variabel	Normal jika <=80 (0), Prehipertensi jika <90 (1), Hipertensi Stadium I jika <100 (2), Hipertensi Stadium II jika >=100 (3)
Frekuensi Nafas	Variabel	Normal jika 12-20 (0), Tidak Normal I jika <12 (1), Tidak Normal II jika >20 (2), Sangat Tidak Normal jika >25 (3)
Gula Darah	Variabel	Normal jika <=180 (0), Tidak Normal jika <=200 (1), Sangat Tidak Normal jika >200 (2),
Romberg Test	Variabel	Negatif (0), Positif (1)
Keterangan Absen	Label	KERJA, SAKIT

3.3 Pengujian Dengan Google Colaboratory

Adapun tahapan pengujian pada *Data Mining* dengan menggunakan Google Colaboratory adalah sebagai berikut :

1. Unggah Data Ke Google Drive

   Data yang disiapkan merupakan *data training* (Tabel 3) dan *data testing* (Tabel 5) yang disimpan dalam format .xlsx. data ini nantinya digunakan sebagai bahan perhitungan analisis yang diunggah ke halaman Google Drive yang nantinya dipergunakan pada Google Colaboratory.

TABEL 5. *Data Testing*

Kode	Keluhan	Obat Yang Dikonsumsi	Suhu Tubuh	Frekuensi Nadi	Tekanan Sistolik	Tekanan Diastolik	Frekuensi Nafas	Gula Darah	Romberg Test	Keterangan
T00001	1	0	1	0	1	1	0	1	0	SAKIT
T00002	0	1	1	1	0	0	2	0	1	KERJA
T00003	0	1	1	1	0	0	1	1	0	SAKIT
T00004	1	1	1	1	2	1	0	2	0	SAKIT
T00005	1	0	1	1	0	0	2	0	1	SAKIT
T00006	0	0	0	0	1	1	0	0	0	KERJA
T00007	0	1	0	0	0	0	3	0	0	KERJA
T00008	1	1	0	1	1	1	1	2	0	KERJA
T00009	1	0	0	0	2	2	2	0	0	SAKIT
T00010	0	1	0	0	0	0	1	1	0	KERJA

Data pada tabel 5 merupakan *data testing* yang digunakan sebagai bahan pengujian, data diatas merupakan contoh kasus absensi karyawan yang ditandai dengan kode T dan diikuti nomor urut data.

2. Unggah Data Ke Google Drive

   Unggah data dilakukan untuk memasukkan data yang akan diuji dalam bentuk format *Excel*. Selanjutnya memasukkan data ke *Google Drive* agar data dapat dipanggil dengan *coding* pada *Google Colaboratory*.

3. Mengimport *Library*

   Pada implementasi *Google Colaboratory* memiliki kelas yaitu kelas proses. Selain itu, perangkat lunak ini menggunakan *library python* yang terdapat pada Gambar 2 berikut :

```
import pandas as pd
import numpy as np
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from google.colab import drive
```

Gambar 2. *Cellcode Import Library*

TABEL 6. Deskripsi *Import Library*

No	Kode	deskripsi
1	import pandas as pd	Untuk menganalisis, membersihkan, mengeksplorasi.
2	import numpy as np	Menyediakan obyek array, menganalisis numerik, melakukan matrix, dan operasi matematika.
3	from sklearn.neighbors import KNeighborsClassifier	Untuk perhitungan metode KNN
4	from sklearn.naive_bayes import GaussianNB	Untuk perhitungan metode NBC
5	from google.colab import drive	Pemanggilan data dari Google Drive

4. Memanggil Data Dari *Google Drive*

Pada tahap ini melakukan pemanggilan data yang sudah diunggah ke *Google Drive* sebelumnya dengan *cellcode* sebagai berikut :

```
drive.mount ('/content/drive/')
data_train = pd.read_excel('/content/drive/MyDrive/data training
PT.SDI.xlsx')
data_test = pd.read_excel('/content/drive/MyDrive/data testing
PT.SDI.xlsx')
data_train.head(10)
data_test.head(10)
```

Gambar 3. *Cellcode* Pemanggilan *Dataset*

Selanjutnya menjalankan *cell code* pada Gambar 3 diatas dan menghasilkan pemanggilan data seperti pada gambar berikut :

```
data_train.head(10)
```

	Kode	Keluhan	Obat Yang Dikonsumsi	Suhu Tubuh	Frekuensi Nadi	Tekanan Sistolik	Tekanan Diastolik	Frekuensi Nafas	Gula Darah	Romberg Test	Keterangan
0	A00001	1	0	1	0	1	1	0	1	0	SAKIT
1	A00002	1	0	1	1	0	0	3	0	1	SAKIT
2	A00003	1	1	1	1	0	0	1	0	0	SAKIT
3	A00004	1	1	1	1	2	2	0	2	0	SAKIT
4	A00005	1	0	1	1	0	0	3	0	1	SAKIT
5	A00006	0	0	0	0	1	1	0	0	0	KERJA
6	A00007	0	1	0	0	0	0	3	0	0	KERJA
7	A00008	0	1	0	1	1	1	1	2	0	KERJA
8	A00009	0	0	0	0	2	2	1	0	0	KERJA
9	A00010	0	0	0	0	0	0	1	0	0	KERJA

Gambar 4. *Data Training*

```
[ ] data_test.head(10)
```

	Kode	Keluhan	Obat Yang Dikonsumsi	Suhu Tubuh	Frekuensi Nadi	Tekanan Sistolik	Tekanan Diastolik	Frekuensi Nafas	Gula Darah	Romberg Test	Keterangan
0	T00001	1	0	1	0	1	1	0	1	0	SAKIT
1	T00002	0	1	1	1	0	0	2	0	1	KERJA
2	T00003	0	1	1	1	0	0	1	1	0	SAKIT
3	T00004	1	1	1	1	2	1	0	2	0	SAKIT
4	T00005	1	0	1	1	0	0	2	0	1	SAKIT
5	T00006	0	0	0	0	1	1	0	0	0	KERJA
6	T00007	0	1	0	0	0	0	3	0	0	KERJA
7	T00008	1	1	0	1	1	1	1	2	0	KERJA
8	T00009	1	0	0	0	2	2	2	0	0	SAKIT
9	T00010	0	1	0	0	0	0	1	1	0	KERJA

Gambar 5. *Data Testing*

5. Membagi *Dataset Training* Dan *Testing* Serta Menentukan Variabel

Dataset pada Gambar 4 dan 5 selanjutnya dilakukan proses pembagian dataset dan penentuan variabel yang digunakan sebagai bahan perhitungan. Berikut adalah tampilan *cell code* dan *output* dari *coding* yang dijalankan :

```
#membagi dataset training dan testing serta menentukan variabel
dtrain = data_train.copy()
dtrain = dtrain.drop(columns=["Kode"], axis=1).to_numpy()
dtest = data_test.copy()
dtest = dtest.drop(columns=["Kode"], axis=1).to_numpy()

X_train = dtrain[:,0:-1]
X_test = dtest[:,0:-1]

y_train = dtrain[:,-1]
y_test = dtest[:,-1]
```

Gambar 6. *Cellcode* Pembagian *Dataset* dan Penentuan Variabel

6. Menjalankan K-Nearest Neighbor Ke Dataset

Setelah dataset telah dibagi dan ditentukan variabelnya selanjutnya melatih *data training* dengan metode KNN dan melakukan pengujian pada *data testing*. Adapun tampilan *cell code* dan *output* dari *coding* yang dijalankan adalah sebagai berikut :

```
knn = KNeighborsClassifier()
knn.fit(X_train,y_train)

hasilPrediksi_knn = knn.predict(X_test)
print("Label Sebenarnya : ", y_test)
print("Hasil Prediksi : ", hasilPrediksi_knn)

prediksiBenar_knn = (hasilPrediksi_knn == y_test).sum()
prediksiSalah_knn = (hasilPrediksi_knn != y_test).sum()
print("Akurasi K-Nearest Neighbor : ",
prediksiBenar_knn/(prediksiBenar_knn+prediksiSalah_knn)*100,"%")
```

Gambar 7. *Cellcode* Perhitungan K-Nearest Neighbor

```
Label Sebenarnya :  ['SAKIT' 'KERJA' 'SAKIT' 'SAKIT' 'SAKIT' 'KERJA' 'KERJA' 'KERJA' 'SAKIT'
 'KERJA']
Hasil Prediksi :  ['SAKIT' 'KERJA' 'KERJA' 'SAKIT' 'SAKIT' 'KERJA' 'KERJA' 'KERJA' 'KERJA'
 'KERJA']
Akurasi K-Nearest Neighbor :  80.0 %
```

Gambar 8. Hasil Klasifikasi Dengan KNN

7. Menjalankan Naïve Bayes Classifier Ke Dataset

Selanjutnya melatih *data training* dengan metode NBC dan melakukan pengujian pada *data testing*. Adapun tampilan *cell code* dan *output* dari *coding* yang dijalankan adalah sebagai berikut :

```
nbc = GaussianNB()
nbc.fit(X_train,y_train)

hasilPrediksi_nbc = nbc.predict(X_test)
print("Label Sebenarnya : ", y_test)
print("Hasil Prediksi : ", hasilPrediksi_nbc)

prediksiBenar_nbc = (hasilPrediksi_nbc == y_test).sum()
prediksiSalah_nbc = (hasilPrediksi_nbc != y_test).sum()
print("Akurasi Naive Bayes : ",
prediksiBenar_nbc/(prediksiBenar_nbc+prediksiSalah_nbc)*100,"%")
```

Gambar 9. *Cellcode* Perhitungan Naïve Bayes Classifier

```
Label Sebenarnya :  ['SAKIT' 'KERJA' 'SAKIT' 'SAKIT' 'SAKIT' 'KERJA' 'KERJA' 'KERJA' 'SAKIT'
 'KERJA']
Hasil Prediksi :  ['SAKIT' 'SAKIT' 'KERJA' 'SAKIT' 'SAKIT' 'KERJA' 'KERJA' 'SAKIT' 'SAKIT'
 'KERJA']
Akurasi Naive Bayes :  70.0 %
```

Gambar 10. Hasil Klasifikasi Dengan NBC

8. Perbandingan Hasil Pergitungan Metode KNN dan NBC

Melakukan perbandingan hasil klasifikasi kedua metode. Adapun *cellcode* dan *output* yang dihasilkan seperti pada gambar dibawah ini :

```
data_df = data_test.copy()
data_df = data_df.drop(['Keluhan', 'Obat Yang Dikonsumsi', 'Suhu Tubuh',
'Frekuensi Nadi', 'Tekanan Sistolik', 'Tekanan Diastolik', 'Frekuensi
Nafas', 'Gula Darah', 'Romberg Test'], axis = 1)
data_df['Pred. KNN'] = hasilPrediksi_knn
data_df['Pred. NBC'] = hasilPrediksi_nbc
data_df
```

Gambar 11. *Cellcode* Perbandingan Hasil Klasifikasi

	Kode	Keterangan	Pred. KNN	Pred. NBC
0	T00001	SAKIT	SAKIT	SAKIT
1	T00002	KERJA	KERJA	SAKIT
2	T00003	SAKIT	KERJA	KERJA
3	T00004	SAKIT	SAKIT	SAKIT
4	T00005	SAKIT	SAKIT	SAKIT
5	T00006	KERJA	KERJA	KERJA
6	T00007	KERJA	KERJA	KERJA
7	T00008	KERJA	KERJA	SAKIT
8	T00009	SAKIT	KERJA	SAKIT
9	T00010	KERJA	KERJA	KERJA

Gambar 12. Perbandingan Hasil Klasifikasi

Pada gambar diatas terdapat data Keterangan (Data Aktual), Hasil Prediksi metode KNN dan Hasil Prediksi metode NBC. Perhitungan akurasi dapat dilakukan dengan perhitungan *Confusion Matrix* sebagai berikut :

TABEL 7. Perhitungan Akurasi Dengan *Confusion Matrix*

KNN	**AKTUAL**		**CONFUSION MATRIX**	
PREDIKSI	SAKIT	KERJA		
SAKIT	3	0	ACCURACY	80%
KERJA	2	5		
NBC	**AKTUAL**		**CONFUSION MATRIX**	
PREDIKSI	SAKIT	KERJA		
SAKIT	4	2	ACCURACY	70%
KERJA	1	3		

Jika dilihat dari hasil perhitungan akurasinya dari kedua metode, metode KNN mendapatkan nilai akurasi sebesar 80% sedangkan metode NBC dengan nilai akurasi sebesar 70%, maka dapat dikatakan berdasarkan dari *data testing* yang ada metode KNN dapat mengklasifikasikan data lebih baik daripada metode NBC.

4. KESIMPULAN

Pada penelitian ini klasifikasi kondisi kesehatan karyawan bagi PT. Serba Dinamik Indonesia berhasil diterapkan dengan data mining menggunakan metode *Naive Bayes Classifier* dan *K-Nearest Neighbor* yang dituangkan kedalam bahasa pemrograman Python (Google Colaboratory) dengan menggunakan data sebanyak 9910 absensi yang digunakan sebagai *data training* dan 10 data yang digunakan sebagai *data testing*. Hasil klasifikasi dari kedua metode menggunakan 10 *data testing* yang diuji menunjukkan bahwa metode *K-Nearest Neighbor* melakukan pengklasifikasian dengan baik dengan tingkat akurasi sebesar 80% sementara metode *Naïve Bayes Classifier* sebesar 70%.

UCAPAN TERIMAKASIH

Tuliskan ucapan terimakasih, contoh penulisannya adalah: Terimakasih kepada pihak pemberi dana penelitian, kepada lembaga/ orang yang membantu penelitian, kepada orang yang membantu dalam diskusi, dll.

DAFTAR PUSTAKA

[1] Cahyanti, F. L. D., Gata, W., & Sarasati, F. (2021). Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit. *Jurnal Ilmiah Universitas Batanghari Jambi*.

[2] Yuliansyah, M. R., Muslimin, B., & Franz, A. (2022). Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Classifier Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda. *Adopsi Teknologi Dan Sistem Informasi (ATASI)*.

[3] Chaerudin, R. A., Mariana, R., & Wadu, B. (2022). Implementasi Algoritma Naïve Bayes Untuk Analisis Klasifikasi Survei Kesehatan Mental (Studi Kasus : Open Sourcing Mental Illness). *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*.

[4] Putri, R. W., Ristyawan, A., & Muzaki, M. N. (2022). Perbandingan Kinerja Algoritma K-NN dan NBC Untuk Klasifikasi Penyakit Jantung. *JTECS : Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem & Komputer*, *2*(2), 143–154.

[5] Senika, A., & Iskandar, D. (2022). Implementasi Metode Naïve Bayes Dalam Penilaian Kinerja Sales Marketing Pada PT. Pachira Distrinusa. *Jurnal Media Informatika Budidarma*, *6*(1), 701–709.

[6] Dina, A., Permana, I., Muttakin, F., & Maita, I. (2023). Perbandingan Algoritma NBC, KNN, dan C4.5 Untuk Klasifikasi Penerima Bantuan Program Keluarga Harapan. *Jurnal Media Informatika Budidarma*.

[7] Akmal, K., Faqih, A., & Dikananda, F. (2023). Perbandingan Metode Algoritma Naïve Bayes Dan K-Nearest Neighbors Untuk Klasifikasi Penyakit Stroke. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 470–477.

[8] Batubara, G. M. C., Desiani, A., & Amran, A. (2023). Klasifikasi Jamur Beracun Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbors. *Jurnal Ilmu Komputer Dan Informatika*.

[9] Maricar, M. A., & Pramana, D. (2019). Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali. *Jurnal Sistem Dan Informatika (JSI)*.

[10] Hidayati, N., Utomo, V. G., & Wakhidah, N. (2019). Penerapan Metode Naïve Bayes Dalam Pemeriksaan Kesehatan. *Jurnal Informatika Upgris*.

[11] Wahyuningsih, U., Sulistiyo, E., Rusjdi, H., Alfalah, W., Sudirmanto, & Prabowo, E. (2021). Pengenalan Keselamatan Dan Kesehatan Kerja Di PT. Cita Rasa Palembang. *Terang: Jurnal Pengabdian Pada Masyarakat Menerangi Negeri*.
[12] RB. Asyim, & Yulianto. (2022). Perilaku Konsumsi Obat Tradisional dalam Upaya Menjaga Kesehatan Masyarakat Bangsawan Sumenep. *Jurnal Keperawatan.*
[13] Pakpahan, M., Siregar, D., Susilawaty, A., Tasnim, M., Ramdany, R., Manurung, E. I., Sianturi, E., Tompunu, M. R. G., Sitanggang, Y. F., & M, M. (2021). Promosi Kesehatan & Prilaku Kesehatan. In Tim Kreatif Kita Menulis (Ed.), *Jakarta: EGC*. Yayasan Kita Menulis.
[14] Pranitasari, D., & Khotimah, K. (2021). Analisis Disiplin Kerja Karyawan pada PT. Bont Technologies Nusantara. *Jurnal Akuntansi Dan Manajemen*, 38.
[15] Rahma, L., Syaputra, H., Mirza, A. H., & Purnamasari, S. D. (2021). Objek Deteksi Makanan Khas Palembang Menggunakan Algoritma YOLO (You Only Look Once). *Jurnal Nasional Ilmu Komputer*.

Prosiding- SEMASTER*: Seminar Nasional Teknologi Informasi & Ilmu Komputer* is licensed under a Creative Commons Attribution International (CC BY-SA 4.0)