

PREDIKSI LEVEL OBESITAS BERDASARKAN KEBIASAAN MAKAN DAN KONDISI FISIK

Abdurrizqo Arrahman¹, Muhammad Fardan Fauzan², Mohammad Aditya Putra³, Prasetya Naufal Rahmandita⁴

Email: 1abdurrizqo@student.ub.ac.id, 2fardanfauzan@student.ub.ac.id, 3maditiyaputra@student.ub.ac.id,
4prasetyanaufal@student.ub.ac.id

Abstrak

Kesehatan merupakan hal yang sangat penting, dalam menjaga kesehatan seseorang harus menjaga kebutuhan makanannya dan pilihan makanannya. Banyak kasus orang yang kelebihan berat badan atau disebut obesitas di dunia. Obesitas dipengaruhi oleh bagaimana seseorang mengatur jadwal makan mereka dan seberapa banyak serta apa saja kandungan dari makanan mereka. Obesitas itu sendiri memiliki beberapa level yang dimana semakin tinggi levelnya akan semakin parah. Dalam memprediksi level obesitas ataupun kondisi berat badan seseorang dapat dilakukan menggunakan teknologi *data mining*, seperti tujuan dibuatnya penelitian ini yaitu memprediksi level obesitas berdasarkan kebiasaan makan dan kondisi fisik seseorang dengan menggunakan algoritma *K-Nearest Neighbor* (KNN) yang diproses menggunakan bahasa pemrograman python. Hasil penelitian memperoleh akurasi menggunakan algoritma *K-Nearest Neighbor* (KNN) sebesar 87% dengan nilai $k = 1$ serta ukuran dataset 90-10.

Kata kunci: *KNN, python, makanan, obesitas, kesehatan*

1. PENDAHULUAN

A. Latar Belakang

Obesitas adalah salah satu masalah yang diabaikan masyarakat pada masa kini. Hal ini adalah sebuah permasalahan yang terjadi di seluruh dunia. Di negara berkembang, obesitas telah menjadi epidemi global di kalangan anak-anak, remaja, dewasa dan orang tua. Prevalensi kegemukan, termasuk obesitas, cenderung meningkat seiring bertambahnya usia dan mencapai puncaknya pada usia dewasa. Status gizi pada orang dewasa di atas usia 18 tahun didominasi oleh masalah obesitas. Menurut Stefanie Eschenbacher dan Carlos Jasso yang dilansir pada laman Reuters (2021) mengatakan bahwa dalam 6 tahun terakhir, warga Meksiko dewasa yang berumur 20 atau lebih mengalami peningkatan empat persen menjadi 75,2%; 39,1% mengalami kelebihan berat badan dan 36.1% lainnya merupakan obesitas.

Dengan perkembangan restoran cepat saji dan perkembangan teknologi, manusia diberi kemudahan untuk mengakses dan memesan makanan dengan sangat mudah hanya dari ponsel

mereka. Dengan kemudahan yang diberikan, membuat manusia pada masa ini menjadi lebih sedikit bergerak yang mengakibatkan kondisi fisik menurun. Juga dengan berkembangnya restoran atau pedagang makanan cepat saji tentunya memberikan kemudahan juga bagi masyarakat, terlebih mereka tidak perlu menunggu terlalu lama untuk mendapatkan makanan yang mereka inginkan menjadikannya sebagai salah satu pilihan bagi mereka untuk mengisi kekosongan perut. Berdasarkan artikel pada laman OmeCare, beberapa alasan mengapa makanan cepat saji menyebabkan obesitas adalah bahan-bahan yang digunakan tidak sehat, porsi makanan yang lebih besar namun didapatkan dengan harga yang murah jika dibandingkan dengan makanan sehat dengan porsi yang sama, serta kemudahan pelayanan yang diberikan sehingga tidak membuang waktu terlalu banyak.

Dengan Uraian diatas, maka diperlukannya analisis prediksi data untuk menentukan level obesitas seseorang berdasarkan kebiasaan makan dan kondisi fisik.

2 Proyek Akhir Mata Kuliah Pengantar Sains Data

B. Tujuan

Ada juga tujuan dari penelitian ini adalah:

1. Untuk memperoleh pengetahuan baru
2. Memprediksi tingkat obesitas seseorang berdasarkan pola makan dan kondisi fisik

C. Identifikasi Masalah

Berdasarkan pemaparan latar belakang diatas, maka dapat diidentifikasi beberapa masalah:

1. Obesitas menjadi permasalahan yang diabaikan masyarakat.
2. Lebih dari separuh warga dewasa di Meksiko mengalami kelebihan berat badan.
3. Kondisi fisik menjadi suatu aspek pertimbangan dalam penyakit obesitas.
4. Pola makan yang tidak baik merupakan suatu aspek penting dalam penyakit obesitas.

D. Referensi Penelitian Lain

Kami memiliki referensi dari penelitian lain yaitu penelitian dari Yulia Kurniawati, Rudi Fakhriadi, dan Fahrini Yulidasari dengan judul Hubungan Antara Pola Makan, Asupan Energi, Aktivitas Fisik, dan Durasi Tidur Dengan Kejadian Obesitas Pada Polisi (2016). Namun ada perbedaan pada penelitian kami dan referensi penelitian yaitu kami menggunakan variabel kebiasaan makan dan kondisi fisik sedangkan referensi dari penelitian kami menggunakan variabel pola makan, asupan energi, aktivitas fisik, dan durasi tidur serta objek penelitiannya adalah para polisi di Kepolisian Resort Kota Banjarmasin, dan pada penelitian kami menggunakan data dari 3 negara yaitu Meksiko, Kolombia, dan Peru.

2. METODOLOGI

Adapun alur proses dari metodologi yang akan kami terapkan adalah sebagai berikut :



Gambar 1. Model Penelitian

2.1. Studi Literatur

Pada tahap ini peneliti harus memahami terlebih dahulu mengenai metode penelitian yang akan digunakan yaitu dengan melihat pada referensi yang didapat dari jurnal, buku maupun penelitian sebelumnya yang menggunakan metode penelitian yang sama. Untuk mendukung teori yang akan digunakan dalam penelitian ini, peneliti akan melakukan beberapa kutipan kalimat yang dilakukan dengan sitasi.

2.2. Mengumpulkan Dataset

Proses pengumpulan dataset dilakukan dengan mencari pada berbagai *platform website* yang menyediakan dataset analisis. pada penelitian ini dilakukan pencarian dataset pada *website* <https://kaggle.com> dan archive.ics.uci.edu. Dalam kedua *website* tersebut terdapat berbagai macam dataset yang dapat digunakan sebagai bahan analisis klasifikasi.

2.3. Mengolah dan Menyiapkan Dataset

Dari berbagai macam dataset yang terdapat pada kaggle.com dan archive.ics.uci.edu kami memilih satu dataset yaitu klasifikasi jenis obesitas berdasarkan pola hidup seseorang. Dataset ini terdiri dari 9 fitur dan 7 tipe kelas analisis.

Karena pada beberapa fitur pada dataset masih bertipe data string maka data akan di encode menjadi tipe data float. setelah berhasil nantinya data yang di

encode akan berubah menjadi 0,1,2 dan seterusnya sebagaimana angka yang mewakilinya.

data yang telah diolah selanjutnya akan disiapkan untuk dilakukan analisa. dari sekitar 2000 data pada dataset tersebut nantinya akan dipisah menjadi dua bagian yaitu data uji dan data analisis. pembagian ini akan dilakukan secara berulang dengan jumlah yang berbeda-beda.

2.4. Melakukan Analisis dengan teori KNN

Seluruh data yang telah dibagi menjadi data latih dan data uji selanjutnya akan diproses dengan algoritma KNN. Untuk mempermudah perhitungan maka digunakan library pada python yaitu *scikit-learn*. dengan melakukan import dan menggunakan method yang telah tersedia proses penghitungan akan dilakukan secara otomatis.

Setelah proses penghitungan selesai tahapan analisis tidak selesai begitu saja. analisis akan terus dilanjutkan dengan berbagai macam kemungkinan. Dalam penelitian ini kami kembali melakukan pemisahan data latih dengan uji dengan berbagai jumlah. selanjutnya kami juga mengubah nilai K dari penghitungan KNN dengan range 1 - 10 pada setiap variasi jumlah data latih. hal ini bertujuan untuk mendapatkan hasil prediksi yang terbaik dan mempermudah melakukan analisis

2.5. Melakukan Evaluasi

Tahapan terakhir, seluruh hasil analisis dari berbagai macam kemungkinan nilai K dan perubahan data uji-latih akan dicatat. seluruh hasil analisis akan dilakukan visualisasi pada diagram agar mudah dalam membaca dan mengevaluasi hasil analisis.

Dari seluruh hasil analisis akan dievaluasi seperti mencari tahu fitur yang memiliki pengaruh terbesar, data yang bermasalah, penyebab hasil prediksi menurun dan nilai prediksi terbaik. hal ini dilakukan untuk memastikan nilai prediksi saat diberikan data baru nantinya.

3. HASIL DAN PEMBAHASAN

Pada bagian ini dituliskan hasil dari proses pengerjaan proyek akhir berikut penjelasan dan pembahasannya. Apabila diperlukan subbab, silakan diikuti saja format penulisan subbab seperti pada Bab 2.

Semua tabel dan gambar harus jelas/tidak kabur/buram. Ukuran huruf pada tabel dan gambar harus dapat dibaca oleh mata normal dengan mudah. Posisi tabel atau gambar di suatu halaman, sebaiknya terletak di bagian atas atau bawah halaman pada tiap kolom. Contoh dapat dilihat pada tabel 1 atau gambar 1. Meletakkan tabel atau gambar ditengah

halaman atau paragraf supaya dihindari. Tabel dan gambar diletakkan pada posisi tengah setiap kolom (*center alignment*).

3.1. Dataset

Dataset pada penelitian ini berasal dari web yang dapat digunakan oleh masyarakat umum, disini kami menggunakan dataset dari UCI Machine Learning Repository dengan judul *Estimation of obesity levels based on eating habits and physical condition Data Set*.

Dataset yang digunakan berasal dari laman berikut:

<https://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition+>

Tabel 1. Atribut dan nilai pada dataset

Atribut	Nilai
Gender	(Male, Female) diisikan jenis kelamin sesuai subjek
Age	Usia disesuaikan dengan data yang didapatkan
Height	Tinggi disesuaikan dengan data yang didapatkan
Weight	Berat disesuaikan dengan data yang didapatkan
Family_history_with_overweight	Memiliki riwayat keluarga yang kegemukan (yes, no)
FAVC (Frequent consumption of high caloric food)	Sering mengonsumsi makanan tinggi kalori (yes, no)
FCVC (Frequency of consumption of vegetables)	Seberapa sering mengonsumsi sayuran (Never, sometimes, always)
NCP (Number of main meals)	Berapa kali anda makan makanan utama (Between 1 y 2, three, more than three)
CAEC (Consumption of food between meals)	Apakah anda makan di sela-sela waktu makan utama (No, sometimes,

4 Proyek Akhir Mata Kuliah Pengantar Sains Data

	frequently, always)
SMOKE	Apakah anda merokok (yes, no)
CH2O (Consumption of water daily)	Seberapa banyak anda meminum air untuk harian (Less than a liter, between 1 and 2L, more than 2L)
SCC (Calories consumption monitoring)	Apakah anda melakukan pemantauan untuk konsumsi total kalori harian (yes, no)
FAF (Physical activity frequency)	Seberapa sering melakukan aktivitas fisik (i do not have, 1 or 2 days, 2 or 4 days, 4 or 5 days)
TUE (Time using technology devices)	Seberapa banyak waktu yang anda gunakan untuk menggunakan alat elektronik (0-2 hours, 3-5 hours, more than 5 hours)
CALC (Consumption of alcohol)	Seberapa sering anda mengkonsumsi alkohol (i do not drink, sometimes, frequently, always)
MTRANS	Kendaraan yang dipakai (Automobile, motorbike, bike, Public transportation, walking)
NObeyesdad	Melakukan kategori berdasarkan <i>mass body index</i> (Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III)

Berikut fitur dari dataset tersebut:

Fitur

1. FAVC (Frequent consumption of high caloric food)

2. FCVC (Frequency of consumption of vegetables)
3. NCP (Number of main meals)
4. CAEC (Consumption of food between meals)
5. CH2O (Consumption of water daily)
6. SCC (Calories consumption monitoring)
7. FAF (Physical activity frequency)
8. TUE (Time using technology devices)
9. CALC (Consumption of alcohol)

Berikut class dari dataset tersebut:

1. Insufficient Weight
2. Normal Weight
3. Overweight Level I
4. Overweight Level II
5. Obesity Type I
6. Obesity Type II
7. Obesity Type III

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Gender                                2111 non-null   object
1   Age                                   2111 non-null   float64
2   Height                                2111 non-null   float64
3   Weight                                2111 non-null   float64
4   family_history_with_overweight       2111 non-null   object
5   FAVC                                  2111 non-null   object
6   FCVC                                  2111 non-null   float64
7   NCP                                   2111 non-null   float64
8   CAEC                                  2111 non-null   object
9   SMOKE                                 2111 non-null   object
10  CH2O                                  2111 non-null   float64
11  SCC                                   2111 non-null   object
12  FAF                                   2111 non-null   float64
13  TUE                                   2111 non-null   float64
14  CALC                                  2111 non-null   object
15  MTRANS                                2111 non-null   object
16  NObeyesdad                           2111 non-null   object
dtypes: float64(8), object(9)
memory usage: 280.5+ KB
```

Gambar 2. Model awal data

3.2. Data Preprocessing

A. Melakukan pengecekan data kosong

Pada tahapan *Preprocessing* ini dilakukan pengecekan data untuk setiap fitur yang tersedia apakah terdapat data yang kosong dari dataset yang obesitas yang sudah diambil dengan memasukkan perintah `isna`.

```
In [7]: df.isna().sum()
Out[7]: Gender      0
Age      0
Height   0
Weight   0
family_history_with_overweight  0
FAVC     0
FCVC     0
NCP      0
CAEC     0
SMOKE    0
CH2O     0
SCC      0
FAF      0
TUE      0
CALC     0
MTRANS   0
NObeyesdad  0
dtype: int64
```

Gambar 3. Mengecek data kosong pada dataset

B. Melakukan pengecekan duplikasi data dari dataset

Setelah melakukan pengecekan data yang kosong, dilakukanlah langkah selanjutnya yaitu pengecekan duplikasi data untuk melihat berapa jumlah data yang terduplikasi.

```
sum(df.duplicated())
24
```

Gambar 4. Mengecek data duplikasi pada dataset

Pada pengecekan data duplikasi didapatkan terdapat 24 data duplikasi pada dataset, karena terdapat hanya 24 data duplikasi, maka dilakukan pembersihan dengan menggunakan fungsi `drop_duplicates()` pada python.

```
df = df.drop_duplicates()
sum(df.duplicated())
0
```

Gambar 5. Melakukan drop data duplikasi dan mengecek kembali data duplikasi pada dataset

C. Feature Scaling

Feature Scaling adalah cara untuk membuat data numerik pada datasetnya memiliki rentang nilai yang sama sehingga tidak ada satu variabel yang mendominasi variabel lainnya.

D. Melakukan konversi dari data kategorikal menjadi numerik atau proses encoding

Pada dataset yang diambil terdapat berbagai macam kategori yang sudah dijelaskan di atas, dari dataset yang sudah dicek seperti yang sudah diperlihatkan di atas yaitu *objek*, menjadi *integer* atau *float*.

```
from sklearn import preprocessing
le = preprocessing.LabelEncoder()
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
0   Gender                                2111 non-null   object
1   Age                                  2111 non-null   float64
2   Height                              2111 non-null   float64
3   Weight                              2111 non-null   float64
4   family_history_with_overweight       2111 non-null   object
5   FAVC                                 2111 non-null   object
6   FCVC                                 2111 non-null   float64
7   NCP                                  2111 non-null   float64
8   CAEC                                 2111 non-null   object
9   SMOKE                               2111 non-null   object
10  CH2O                                2111 non-null   float64
11  SCC                                  2111 non-null   object
12  FAF                                  2111 non-null   float64
13  TUE                                  2111 non-null   float64
14  CALC                                 2111 non-null   object
15  MTRANS                               2111 non-null   object
16  NObeyesdad                           2111 non-null   object
dtypes: float64(8), object(9)
memory usage: 280.5+ KB
```

Gambar 6. Tipe data pada dataset sebelum dilakukan encoding

Gambar diatas adalah info tipe data pada dataset sebelum dilakukannya encoding, terlihat bahwa terdapat tipe data object dan float. Data dengan tipe data object akan diubah ke tipe data numerik, integer ataupun float. Berikut proses encoding yang dilakukan.

```
Gender=le.fit_transform(df['Gender'])
Age=df['Age']
Height=df['Height']
Weight=df['Weight']
Family_history=le.fit_transform(df['family_history_with_overweight'])
FAVC=le.fit_transform(df['FAVC'])
FCVC=df['FCVC']
NCP=df['NCP']
CAEC=le.fit_transform(df['CAEC'])
SMOKE=le.fit_transform(df['SMOKE'])
CH2O=df['CH2O']
SCC=le.fit_transform(df['SCC'])
FAF=df['FAF']
TUE=df['TUE']
CALC=le.fit_transform(df['CALC'])
MTRANS=le.fit_transform(df['MTRANS'])
NObeyesdad=le.fit_transform(df['NObeyesdad'])
```

Gambar 7. Proses encoding 1

Setiap atribut pada dataset dengan tipe data object diubah menjadi integer atau float.

6 Proyek Akhir Mata Kuliah Pengantar Sains Data

```
data={
  "Gender":Gender,
  "Age":Age,
  "Height":Height,
  "Weight":Weight,
  "family_history":Family_history,
  "FAVC":FAVC,
  "FCVC":FCVC,
  "NCP":NCP,
  "CAEC":CAEC,
  "SMOKE":SMOKE,
  "CH20":CH20,
  "SCC":SCC,
  "FAF":FAF,
  "TUE":TUE,
  "CALC":CALC,
  "MTRANS":MTRANS,
  "NObeyesdad":NObeyesdad
}
```

```
dataset=pd.DataFrame(data)
```

Gambar 8. Proses encoding 2

Setelah dilakukan proses encoding, maka data yang sebelumnya bertipe object akan berubah menjadi integer atau float dan data yang telah diubah dimasukkan ke dataframe dan dikembalikan menjadi dataset.

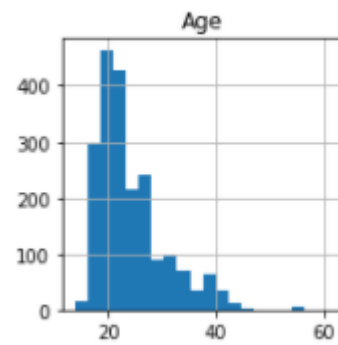
3.3. Data Exploration

Langkah ini dilakukan untuk mendapatkan informasi - informasi dan memahami data set tersebut. *Data Exploration* dilakukan untuk dengan menampilkan berbagai informasi atribut dalam bentuk grafik.

df.describe()								
	Age	Height	Weight	FCVC	NCP	CH20	FAF	TUE
count	2087.000000	2087.000000	2087.000000	2087.000000	2087.000000	2087.000000	2087.000000	2087.000000
mean	24.353090	1.702674	86.858730	2.421466	2.701179	2.004749	1.012812	0.663035
std	6.368801	0.093186	26.190847	0.534737	0.764614	0.608284	0.853475	0.608153
min	14.000000	1.450000	39.000000	1.000000	1.000000	1.000000	0.000000	0.000000
25%	19.915937	1.630178	66.000000	2.000000	2.697467	1.590922	0.124505	0.000000
50%	22.847618	1.701584	83.101100	2.396265	3.000000	2.000000	1.000000	0.630866
75%	26.000000	1.769491	108.015907	3.000000	3.000000	2.466193	1.678102	1.000000
max	61.000000	1.980000	173.000000	3.000000	4.000000	3.000000	3.000000	2.000000

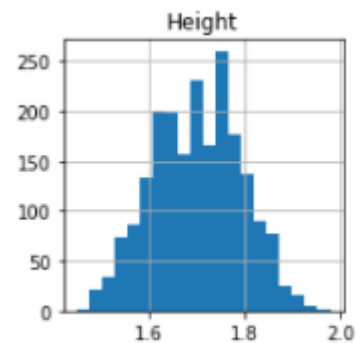
Gambar 9. Informasi dataset

Gambar diatas merupakan informasi dataset, seperti max, min, mean, dan lainnya.



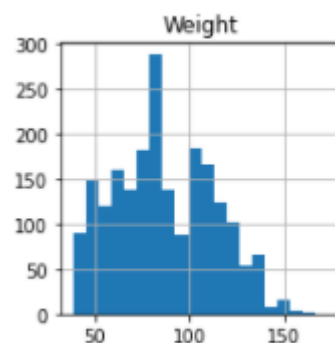
Gambar 10. Persebaran data atribut Age

Gambar diatas merupakan grafik persebaran data untuk atribut Age. Dari grafik tersebut dapat dilihat bahwa paling banyak data dengan umur 20 tahun.



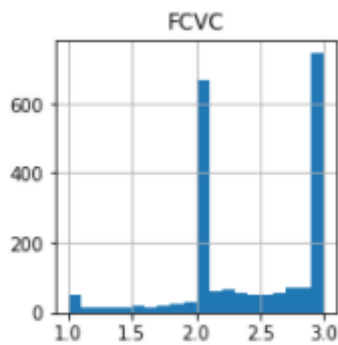
Gambar 11. Persebaran data atribut Height

Gambar diatas merupakan grafik persebaran data untuk atribut Height.



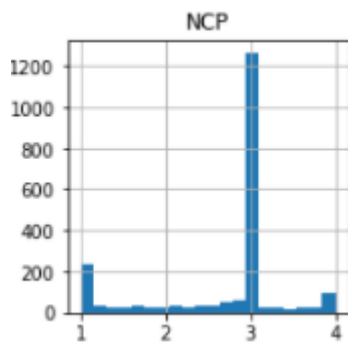
Gambar 12. Persebaran data atribut Weight

Gambar diatas merupakan grafik persebaran data atribut Weight.



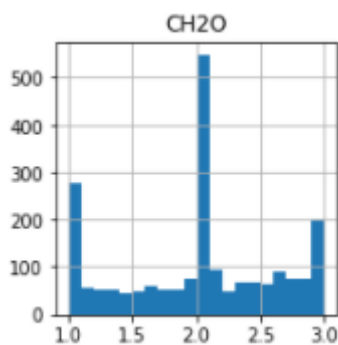
Gambar 13. Persebaran data atribut FCVC

Gambar diatas merupakan grafik persebaran data untuk atribut FCVC.



Gambar 14. Persebaran data atribut NCP

Gambar diatas merupakan grafik persebaran data untuk atribut NCP.



Gambar 15. Persebaran data atribut CH2O

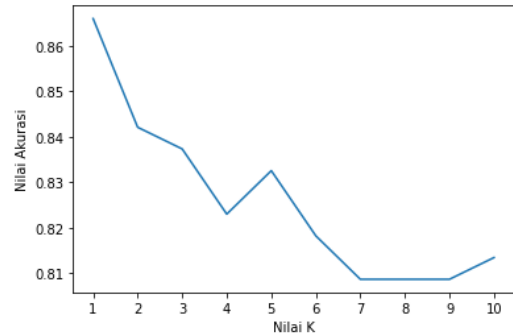
Gambar diatas merupakan grafik persebaran data untuk atribut CH2O. Dari grafik - grafik diatas dapat disimpulkan bahwa tidak ada data outlier dan data - data tersebut termasuk normal.

3.4. Klasifikasi *K-Nearest Neighbor*

Pada tahap ini dilakukan klasifikasi dengan menggunakan algoritma *K-Nearest Neighbor* dengan melakukan analisis dengan ukuran dataset yang berbeda.

Untuk mendapatkan akurasi yang maksimal, maka dilakukan pengecekan dengan mengubah nilai k setiap dilakukannya klasifikasi dan juga dilakukan perubahan penggunaan ukuran dataset. Penggunaan ukuran datasetnya adalah 90-10, 80-20, 70-30, 60-40 dan 50-50, sehingga diperoleh data sebagai berikut:

A. Ukuran Dataset 90-10



Gambar 16. Mengecek nilai $k=1$ sampai $k=10$ dengan ukuran dataset 90-10

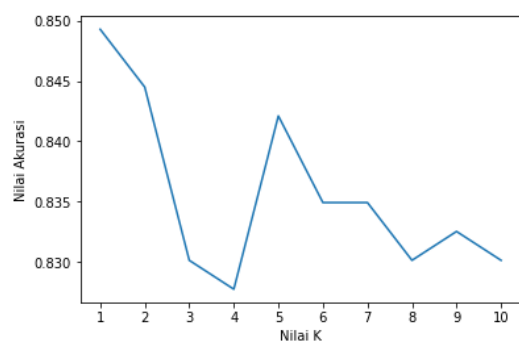
Dari grafik diatas, dapat dilihat bahwa grafik menurun seiring dengan naiknya nilai k , sehingga dapat disimpulkan bahwa nilai $k=1$ memberikan nilai akurasi yang paling tinggi. Berikut detail nilai akurasi dengan $k=1$.

	precision	recall	f1-score	support
0	0.83	0.86	0.84	22
1	0.74	0.61	0.67	28
2	0.93	0.88	0.90	42
3	0.94	1.00	0.97	30
4	0.97	1.00	0.98	30
5	0.88	0.85	0.87	27
6	0.74	0.83	0.78	30
accuracy			0.87	209
macro avg	0.86	0.86	0.86	209
weighted avg	0.87	0.87	0.86	209

Akurasi = 0.8660287081339713

Gambar 17. Nilai akurasi dengan $k=1$ pada ukuran dataset 90-10

B. Ukuran Dataset 80-20



Gambar 18. Mengecek nilai $k=1$ sampai $k=10$ dengan ukuran dataset 80-20

Dari grafik diatas, dapat dilihat bahwa grafik juga menurun seiring dengan naiknya

8 Proyek Akhir Mata Kuliah Pengantar Sains Data

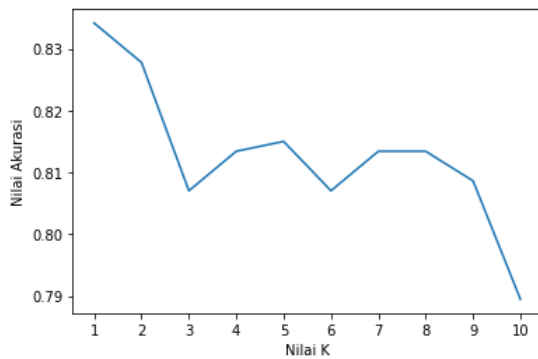
nilai k, sehingga dapat disimpulkan bahwa nilai k = 1 memberikan nilai akurasi yang paling tinggi. Berikut detail nilai akurasi dengan k = 1.

	precision	recall	f1-score	support
0	0.82	0.89	0.85	46
1	0.71	0.57	0.63	51
2	0.90	0.87	0.88	75
3	0.97	0.96	0.96	70
4	0.97	1.00	0.98	60
5	0.78	0.77	0.77	56
6	0.72	0.83	0.78	60
accuracy			0.85	418
macro avg	0.84	0.84	0.84	418
weighted avg	0.85	0.85	0.85	418

Akurasi = 0.8492822966507177

Gambar 19. Nilai akurasi dengan k = 1 pada ukuran dataset 80-20

C. Ukuran Dataset 70-30



Gambar 20. Mengecek nilai k=1 sampai k = 10 dengan ukuran dataset 70-30

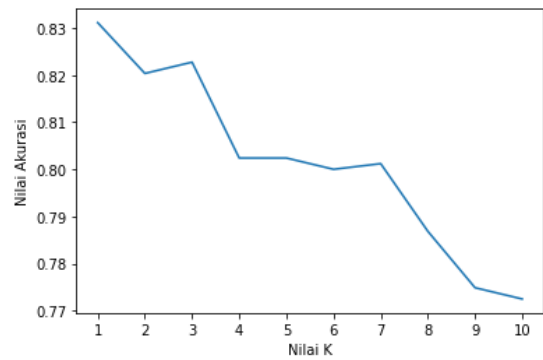
Dari grafik diatas, dapat dilihat bahwa grafik menurun seiring dengan naiknya nilai k, sehingga dapat disimpulkan bahwa nilai k = 1 memberikan nilai akurasi yang paling tinggi. Berikut detail nilai akurasi dengan k = 1.

	precision	recall	f1-score	support
0	0.79	0.92	0.85	79
1	0.71	0.51	0.59	77
2	0.88	0.88	0.88	113
3	0.98	0.93	0.95	98
4	0.98	1.00	0.99	92
5	0.71	0.71	0.71	77
6	0.72	0.81	0.76	91
accuracy			0.83	627
macro avg	0.82	0.82	0.82	627
weighted avg	0.83	0.83	0.83	627

Akurasi = 0.8341307814992025

Gambar 21. Nilai akurasi dengan k = 1 pada ukuran dataset 70-30

D. Ukuran Dataset 60-40



Gambar 22. Mengecek nilai k=1 sampai k = 10 dengan ukuran dataset 60-40

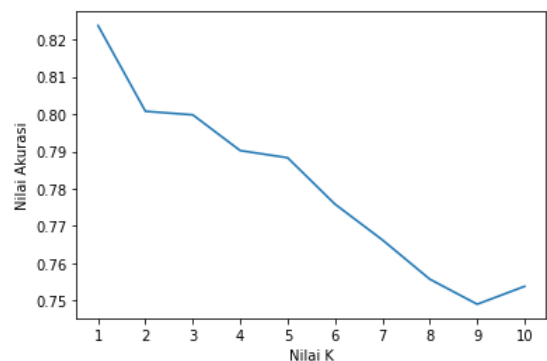
Dari grafik diatas, dapat dilihat bahwa grafik juga menurun seiring dengan naiknya nilai k, sehingga dapat disimpulkan bahwa nilai k = 1 memberikan nilai akurasi yang paling tinggi. Berikut detail nilai akurasi dengan k = 1.

	precision	recall	f1-score	support
0	0.80	0.93	0.86	119
1	0.65	0.47	0.54	98
2	0.88	0.86	0.87	147
3	0.96	0.93	0.94	120
4	0.98	1.00	0.99	120
5	0.74	0.72	0.73	105
6	0.74	0.82	0.77	126
accuracy			0.83	835
macro avg	0.82	0.82	0.82	835
weighted avg	0.83	0.83	0.83	835

Akurasi = 0.8311377245508982

Gambar 23. Nilai akurasi dengan k = 1 pada ukuran dataset 60-40

E. Ukuran Dataset 50-50



Gambar 24. Mengecek nilai k=1 sampai k = 10 dengan ukuran dataset 50-50

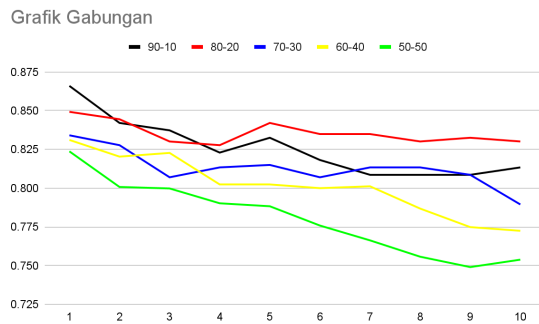
Dari grafik diatas, dapat dilihat bahwa grafik menurun seiring dengan naiknya nilai k, sehingga dapat disimpulkan bahwa nilai k = 1 memberikan nilai akurasi yang paling tinggi. Berikut detail nilai akurasi dengan k = 1.

	precision	recall	f1-score	support
0	0.79	0.90	0.84	141
1	0.63	0.48	0.54	130
2	0.83	0.86	0.85	180
3	0.95	0.93	0.94	152
4	0.99	1.00	0.99	158
5	0.72	0.72	0.72	129
6	0.78	0.81	0.79	154
accuracy			0.82	1044
macro avg	0.81	0.81	0.81	1044
weighted avg	0.82	0.82	0.82	1044

Akurasi = 0.8237547892720306

Gambar 25. Nilai akurasi dengan k = 1 pada ukuran dataset 50-50

Dari hasil - hasil diatas, terlihat bahwa dataset dengan ukuran 90-10 memiliki akurasi yang paling tinggi dengan nilai k = 1, yaitu sebesar 0.87. Berikut grafik gabungan yang dapat digunakan untuk mengkonfirmasi hasil yang telah didapatkan.



Gambar 26. Grafik Gabungan

Dari grafik tersebut terlihat bahwa yang memiliki nilai paling tinggi adalah dataset dengan ukuran 90-10, sedangkan yang paling rendah adalah 50-50. Dengan nilai k = 1 adalah nilai k dengan akurasi paling tinggi di setiap ukuran dataset.

4. PENUTUP

4.1. Kesimpulan

Dari penelitian yang dilakukan dengan menggunakan ukuran dataset yang berubah - ubah dan setiap ukuran dicari nilai k dengan hasil akurasi yang paling tinggi, didapatkan bahwa menggunakan ukuran dataset 90-10 dengan nilai k = 1 memberikan hasil akurasi yang paling tinggi dengan hasil akurasi sebesar 87%.

4.1. Saran

Hasil yang didapatkan tergolong cukup tinggi dan penelitian ini dapat membuka peluang penelitian lain untuk menggunakan algoritma - algoritma lain untuk membandingkan hasil yang didapatkan, apakah dengan algoritma lain

mendapatkan hasil yang lebih bagus atau menjadi lebih buruk.

5. KONTRIBUTOR

Nama	Kontribusi (Tugas)
Abdurrizqo Arrahman	Pencarian data set, pencarian algoritma awal KNN untuk python, melakukan encoding data, perubahan nilai dari kategorial menjadi angka, penyusunan laporan
Muhammad Fardan Fauzan	Mencari referensi jurnal, mencari referensi dataset, penyusunan laporan
Mohammad Aditya Putra	Mencari referensi dataset, mengecek data duplikasi, pencarian algoritma untuk membuat grafik, pembuatan grafik garis, pembagian nilai k, penyusunan laporan
Prasetya Naufal Rahmandita	Mencari referensi dataset, melakukan pengecekan data duplikasi, pencarian algoritma untuk membuat grafik, pembagian ukuran dataset, penyusunan laporan

DAFTAR RUJUKAN

- KUKUH, R., 2018. Melakukan feature scaling pada dataset, [online]: Tersedia di: <<https://medium.com/machine-learning-id/melakukan-feature-scaling-pada-dataset-229531bb08de>> [Diakses 28 November 2021]
- PRASETYOWATI, Y.G., 2019. Klasifikasi menggunakan metode KNN (K-Nearest Neighbor) dalam Python, [online]: Tersedia di: <<https://medium.com/@16611130/klasifikasi-menggunakan-metode-knn-k-nearest-neighbor-dalam-python-a40e79a74101>> [Diakses 28 November 2021]
- Stefanie Eschenbacher and Carlos Jasso., 2021. Factbox: Mexico's obesity epidemic [online]: Tersedia di: <<https://www.reuters.com/world/americas/mexicos-obesity-epidemic-2021-06-16/>> [Diakses 30 November 2021]
- Joos Korstanje., 2021. The k-Nearest Neighbors (kNN) Algorithm in Python [online]: Tersedia di: <<https://realpython.com/knn-python/>> [Diakses 19 November 2021]
- Rahmawati, A., Yulianti, I., Yuliani, Y., Nurhadianto, N., & Novitasari, H. B. (2020). Analisis Algoritma KNN Berbasis Feature Selection untuk Memprediksi Nasabah Pengguna Deposito Melalui Pemasaran Langsung. Swabumi.
- Kurniawati, Y., Fakhriadi, R., & Yulidasari, F. (2016). Hubungan Antara Pola Makan, Asupan Energi, Aktivitas Fisik, dan Durasi Tidur Dengan Kejadian Obesitas Pada Polisi. Jurnal Publikasi Kesehatan Masyarakat Indonesia, 3(3), 112–117.
- Palechor, F. M., & Manotas, A. de la H. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data in Brief, 25, 104344. <https://doi.org/10.1016/j.dib.2019.104344>