

Tugas Analisis KNN dengan dataset Heart Attack

Disusun untuk Memenuhi Tugas Mata Kuliah Pengenalan Pola



Disusun Oleh :

Adinda Aulia Sari (2117051018)

Dosen Pengampu :

Rico Andrian, S.SI., M.Kom

**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG**

2023

DAFTAR ISI

DAFTAR ISI.....	2
BAB I.....	3
1.1 Abstrak	3
1.2 Latar Belakang.....	3
1.3 Rumusan Masalah	3
1.4 Tujuan.....	3
BAB II	4
2.1 Mencari Nilai Akurasi.....	4
2.2 Mencari Nilai K.....	5
BAB III.....	8
3.1 Kesimpulan.....	8
DAFTAR PUSTAKA	8

BAB I

1.1 Abstrak

Algoritma K-Nearest Neighbors (KNN) merupakan metode klasifikasi yang digunakan untuk menganalisis dan memodelkan pola-pola data. Dalam konteks Heart Attack, KNN dapat diterapkan untuk memprediksi risiko serangan jantung berdasarkan atribut-atribut tertentu seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, dan faktor-faktor risiko lainnya. Algoritma KNN bekerja dengan mendekati suatu titik data dengan mengamati label-label dari tetangga terdekatnya.

1.2 Latar Belakang

Serangan jantung adalah penyakit serius yang memerlukan perhatian medis segera. Mengidentifikasi faktor risiko dan mengembangkan model prediksi untuk menentukan risiko serangan jantung dapat menjadi alat yang berharga dalam upaya mencegah dan mengelola penyakit kardiovaskular. Dalam konteks ini, penggunaan algoritma K-Nearest Neighbors (KNN) tampaknya menjadi pendekatan yang menarik dan potensial.

KNN merupakan algoritma pembelajaran mesin yang menggunakan konsep bahwa data dengan properti serupa cenderung memiliki label serupa. Dengan menerapkan KNN pada dataset Serangan Jantung, kita dapat menggunakan informasi dari pasien yang didiagnosis menderita serangan jantung untuk membuat model yang dapat memprediksi risiko serangan jantung pada pasien.

Hal ini membuka peluang untuk menerapkan solusi pencegahan yang lebih efektif dan menangani masalah kesehatan secara lebih proaktif. Penggunaan algoritma KNN dalam konteks serangan jantung menawarkan keuntungan karena mampu menangani hubungan nonlinier antar variabel dan mampu beradaptasi secara fleksibel terhadap pola data yang kompleks. Melalui penelitian ini, kami berharap dapat menemukan pola unik terkait risiko serangan jantung dan merancang model KNN yang dapat diandalkan sebagai alat prediksi.

1.3 Rumusan Masalah

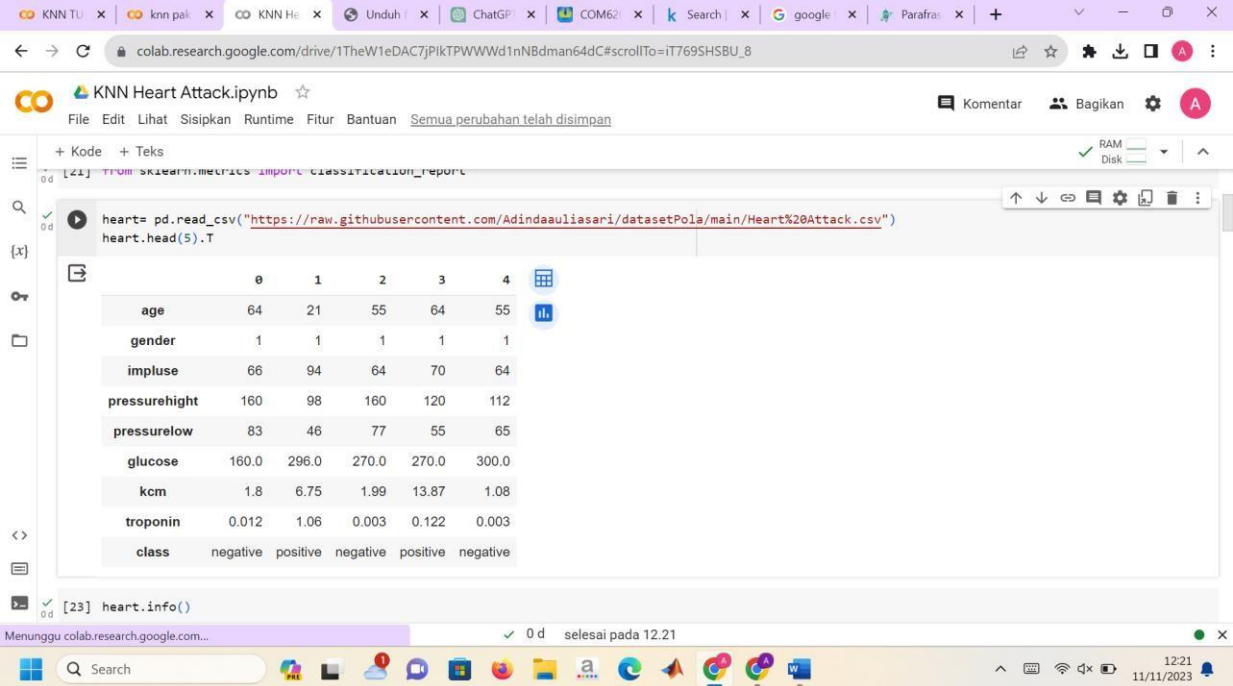
1. Sejauh mana model K-Nearest Neighbor (KNN) efektif digunakan dalam memprediksi risiko diabetes pada dataset Pima Indian, terutama dalam aspek nilai akurasi?
2. Bagaimana cara menentukan nilai K yang optimal pada model KNN untuk mencapai akurasi tertinggi pada dataset ini?

1.4 Tujuan

1. Menganalisis efektivitas model K-Nearest Neighbor (KNN) dalam memprediksi risiko diabetes pada dataset Pima Indian, khususnya dalam aspek nilai akurasi.
2. Menentukan nilai K yang optimal pada model KNN untuk mencapai akurasi tertinggi pada dataset ini.

BAB II

2.1 Mencari Nilai Akurasi



```
[41]: from sklearn.metrics import classification_report

heart= pd.read_csv("https://raw.githubusercontent.com/Adindaauliasari/datasetPola/main/Heart%20Attack.csv")
heart.head(5).T
```

	0	1	2	3	4
age	64	21	55	64	55
gender	1	1	1	1	1
impluse	66	94	64	70	64
pressurehigh	160	98	160	120	112
pressurelow	83	46	77	55	65
glucose	160.0	296.0	270.0	270.0	300.0
kcm	1.8	6.75	1.99	13.87	1.08
troponin	0.012	1.06	0.003	0.122	0.003
class	negative	positive	negative	positive	negative

```
[23]: heart.info()
```

Pertama yang pasti kita perlu mengakses datasetnya, dataset yang saya pakai kali ini adalah dataset heart attack. Dataset ini menampilkan beragam atribut seperti usia, jenis kelamin, denyut nadi, tekanan darah tinggi, tekanan darah rendah, gula darah, KCM, kadar troponin dan kelas positif dan kelas negatif. kumpulan data ini menentukan apakah pasien terkena serangan jantung atau tidak.

Data ini menganalisis serangan jantung dengan 2 class 'positif' atau 'negatif'. artinya Nilai "positif" dapat menunjukkan pasien yang positif terkena serangan jantung, sedangkan "negatif" menunjukkan pasien yang negatif terkena serangan jantung.

```

[24] heart.columns

Preprocessing

df = heart
df['gender'] = df['gender'].apply(lambda x: heart['gender'].unique().tolist().index(x))

[26] df = heart.drop('gender', axis=1)

[27] x = df.drop('class', axis=1)
     y = df['class']

[28] X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 42)

print(X_train.shape)
print(X_test.shape)

# creating list of K for KNN

```

```

[118] print(confusion_matrix(y_test, y_pred))

[[ 54  47]
 [ 31 132]]

print(classification_report(y_test, y_pred))

              precision    recall  f1-score   support

negative     0.64       0.53       0.58        101
positive     0.74       0.81       0.77        163

accuracy     0.70
macro avg     0.69       0.67       0.68        264
weighted avg  0.70       0.70       0.70        264

[120] KnnModel.score(X_test_scaled, y_test)

0.7045454545454546

```

Disini saya melakukan preprocessing serta split untuk data training dan testing yang mana 42% adalah data training dan 20% adalah data testing. Setelah dicari lalu saya mendapatkan nilai akurasi sebesar 70% .

2.2 Mencari Nilai K

Secara umum, formula untuk menghitung jarak antara dua titik dalam ruang N-dimensi menggunakan Euclidean distance adalah:

$$Euclidean Distance = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$$

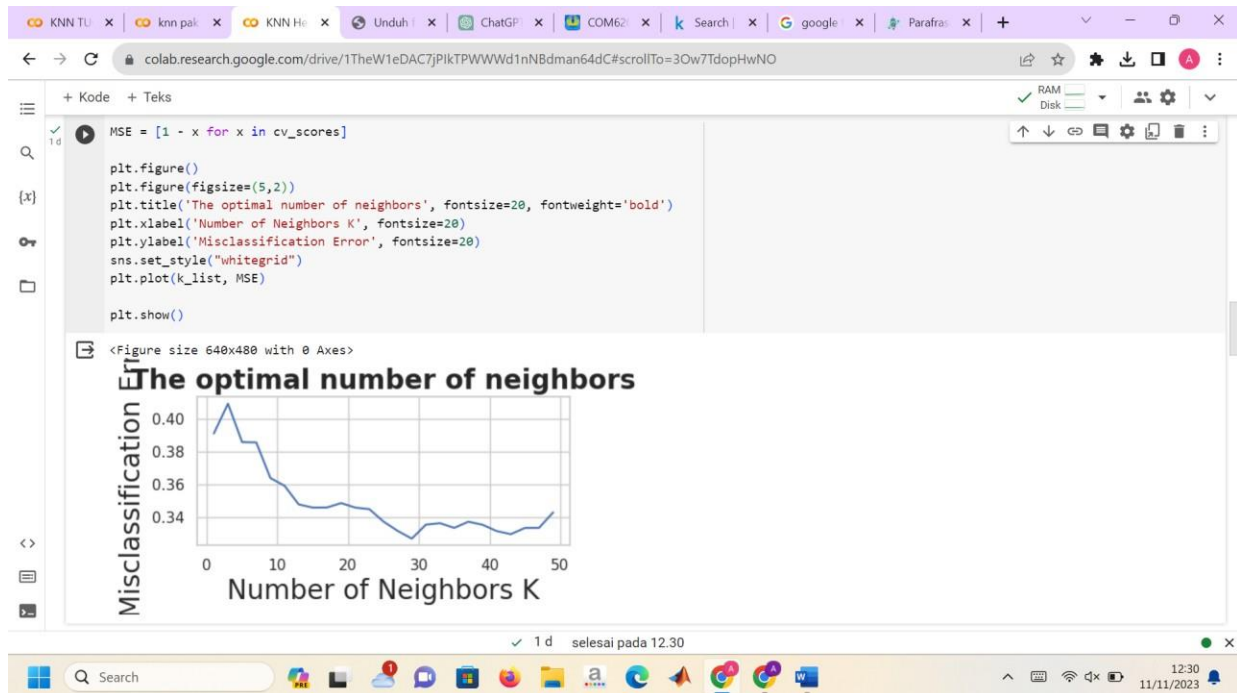
Disini x_i dan y_i adalah koordinat dari 2 titik dimensi i dan N adalah jumlah dimensi.



```
[30] # creating list of K for KNN
      k_list = list(range(1,50,2))
      cv_scores = []

      for k in k_list:
          knn = KNeighborsClassifier(n_neighbors=k)
          scores = cross_val_score(knn, X_train, y_train, cv=20, scoring='accuracy')
          cv_scores.append(scores.mean())
```

Disini saya melakukan test untuk setiap nilai K dari range (1 ,50,2)



Lalu membuat graifk yang menunjukan bahwa di nilai K ke 29 yang memiliki jarak terdekat (tetangga terdekat) antara x_1 dan y_1 (data testing dan data training.

Pemilihan model

```
KnnModel = KNeighborsClassifier(n_neighbors=29)
```

```
[116] KnnModel.fit(X_train_scaled, y_train)
```

KNeighborsClassifier

KNeighborsClassifier(n_neighbors=29)

BAB III

3.1 Kesimpulan

Berdasarkan hasil yang saya dapatkan, penggunaan model K-Nearest Neighbor (KNN) dalam memprediksi dataset Heart Attack menghasilkan tingkat akurasi sebesar 70%. Hal ini menunjukkan bahwa model KNN dapat memberikan prediksi dengan tingkat ketepatan yang baik dalam konteks Heart Attack.

Selain itu, hasil yang saya dapatkan menunjukkan bahwa nilai K yang paling optimal untuk mencapai akurasi tertinggi adalah 29. Pemilihan nilai K yang tepat merupakan langkah kritis dalam meningkatkan kinerja model KNN, dan nilai 29 pada kasus ini terbukti memberikan hasil yang optimal.

DAFTAR PUSTAKA

1. C. Science, “Prediksi Penyakit Jantung dengan Algoritma Klasifikasi,” vol. 5, no. 1, pp. 978–979, 2019.
2. Hasran, “Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor,” *Indones. J. Data Sci.*, vol. 1, no. 1, pp. 1–4, 2020.
3. Lestari M. Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) untuk Mendeteksi Penyakit Jantung. *Fact. Exacta*. 2014; 7(4): 366–371.