**ISSN: 2962-3545**
**Prosiding SAINTEK**: Sains dan Teknologi Vol.2 No.1 Tahun 2023
*Call for papers* dan Seminar Nasional Sains dan Teknologi Ke-2 2023
Fakultas Teknik, Universitas Pelita Bangsa, Februari 2023

# Perbandingan Kinerja Algoritma *Naïve Bayes* Dan *K-Nearest Neighbors* (KNN) Dalam Memprediksi Penyakit Tiroid

## *Comparison Of The Performance Naïve Bayes And K-Nearest Neighbors (KNN) Algorithms In Predicting Thyroid Disease*

**Syarifah Jasmine Putri[1], Adetia Pratama[2], Kevin Attaqwa[3], Rahmaddeni[4]**
[1,2,3,4]Stmik Amik Riau Pekanbaru
[1]2110031806063@sar.ac.id, [2]2110031802143@sar.ac.id, [3]2110031802124@sar.ac.id,
[4]rahmaddeni@sar.ac.id

***Abstract***

*There are still many people who are too busy to forget that greatness is an important thing to maintain, especially the health of the thyroid gland. Thyroid which is a disease with glandular disorders in the form of small organs located in the front of the neck which can cause changes in shape or hormones to become less or too much so that this disease can end in death by using data mining and machine learning algorithms, namely Naïve Bayes and K-Nearest Neighbors (KNN) which is an algorithm with a process that is not too complicated and the process is fast so that researchers can solve problems using thyroid disease data. By using the python framework to find a comparison of the accuracy of the two algorithms, Naïve Bayes algorithm with 63% testing data and 64% training data while the resulting accuracy in the K-Nearest Neighbors (KNN) algorithm is 63%.*

***Keywords***: *Accuracy, Machine Learning, Naïve Bayes, Python, K-Nearest Neighbors (KNN), Thyroid*

**Abstrak**

Masih banyak masyarakat yang terlalu sibuk sehingga melupakan kehesatan merupakan suatu hal yang penting untuk dijaga khusus nya kesehatan pada kelenjar tiroid. Tiroid yang merupakan suatu penyakit dengan gangguan kelenjar yang berupa organ kecil terletak di bagian depan leher yang dapat menyebabkan perubahan pada bentuk atau hormon menjadi lebih sedikit atau terlalu banyak sehingga penyakit ini bisa berujung dengan kematian, dengan menggugunakan algoritma *data mining* dan *machine learning* yaitu *Naïve Bayes* dan *K-Nearest Neighbors (KNN)* yang merupakan salah satu algoritma dengan proses yang tidak terlalu rumit dan proses yang cepat sehingga peneliti dapat memecahkan permasalahan dengan menggunakan data penyakit tiroid. Dengan menggunakan *freamwork python* untuk mencari perbandingan sebuah *accuracy* pada kedua algoritma, algoritma *Naïve Bayes* dengan data testing 63% dan data training 64% sedangkan akurasi yang di hasilkan pada algoritma *K-Nearest Neighbors* (KNN) adalah 63%.

**Kata kunci**: *Accuracy, Machine Learning, Naïve Bayes, Python, K-Nearest Neighbors (KNN),* Tiroid

## Pendahuluan

Masalah kesehatan yang disebabkan oleh adanya gangguan kelenjar berupa organ kecil yang terletak di bagian depan leher serta mengakibatkan perubahan bentuk dan fungsi produksi hormon menjadi lebih berkurang atau terlalu banyak merupakan suatu penyakit yaitu tiroid. Penyakit ini berupa kelenjar endokrin yang dapat membahayakan kesehatan pada tubuh manusia hingga berujung kematian.

Terkadang banyak orang yang terlalu sibuk dengan kepentinga masing-masing sehingga mengabaikan hal yang penting untuk diperhatikan, misalnya kesehatan khususnya kesehatan pada kelenjar tiroid. Dari data yang diperoleh dari Dinas Kesehatan Provinsi Bengkulu, jumlah penderita penyakit tiroid mencapai 2.498 dari 1.249.238 penduduk Provinsi Bengkulu rata- rata dengan usia diatas 15 tahun per Juli 2015.(Aprizum Putra ZM, 2017)

Menggunakan *data mining* dan *machine learning* dengan menggunakan salah satu model klasifikasi sebagai sistem diagnosa penyakit tiroid dapat menjadi suatu alternatif yang tepat. Banyak penelitian yang menggunakan metode *Naïve Bayes* sebagai metode analisa didunia kesehatan karna salah satu kelebihan *Naïve Bayes* adalah impletasi tidak terlalu rumit, proses yang cepat dan memiliki iterasi sedangkan algoritma *K-Nearest Neighbors* atau dapat disingkat KNN memiliki salah satu kelebihan yaitu dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan dengan data yang cukup besar.

Dari penelitian yang dilakukan (Bambang Wijonarko, 2018) yang berkaitan dengan Perbandingan Algoritma *Data Mining Naïve Bayes* Dan *Bayes Network* Untuk Mengidentifikasi Penyakit Tiroid dengan data paseien yang digunakan sebanyak 3711 mengkasilkan akurasi berupa 91.803% menggunakan kurva ROC.

Selain itu penelitian yang di lakukan oleh (Leli Safitri,2022) dengan judul Perbandingan Metode Algoritma *Decision Tree C4.5* Dan *Naïve Bayes* Untuk Memprediksi Penyakit Tiroid dengan data paseien sebanyak 3711 menghasilkan hasil akurasi sebesar 76,02% menggunakan aplikasi *Rapid miner*.

Sedangkan dari penelitian yang dilakukan (Ana Mariyam Puspitasar, 2018) dengan judul Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode *Support Vector Machine* dengan data 122 sehingga memiliki akurasi sebanyak 93.329% untuk semua kelas sehingga terdapat sekitar 30 data untuk setiap kelas dengan jumlah parameter sebanyak 16 data.

Berdasarkan latar belakang tersebut peneliti menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbors (KNN)* dimana metode ini dapat menetukan apakah seseorang terkena penyakit tiroid atau tidak dengan menghitung probabilitas serta kemungkinan dari penyakit dan gejala-gejala yang timbul berdasarkan nilai yang diberikan oleh pakar. Dengan penerapan metode *Naïve Bayes* dan *K-Nearest Neighbors (KNN)* diharapkan dapat mengetahui gejala penyakit tiroid dengan akurat.

## Metode Penelitian

Metode adalah suatu kegiatan yang berkaitan dengan cara kerja (sistematis) yang dapat dipahami pada suatu objek ataupun subjek untuk menemukan suatu jawaban yang dapat dipertanggung jawabkan secara ilmiah. Penelitian adalah suatu kegiatan penulis dalam menganalisa serta kontruksi yang dilakukan secara sistematis, metodologis, dan konsisten yang bertujuan untuk mengetahui hasil yang sedang dikerjakan.

*Machine learning* merupakan pembelajaran dasar pengembangan algoritma komputer untuk mengubah data aksi secara singkat atau dapat juga diartikan sebagai proses perubahan data menjadi sebuah infromasi, peneliti menggunakan program *python* agar dapat memudahkan proses untuk mencari suatu hasil akurasi (Adi Supriyatna, 2018)

Peneliti menggunakan pengembangan algoritma *Naïve Bayes* yang merupakan algoritma yang sederhana berdasarkan penerapan teorema bayes dengan asumsi anatara variable bebas atau independen tidak hanya itu metode ini memanfaatkan probalitas dan statistik. Berikut adalah rumus naïve bayes (Ernawati, 2017)

$$P(A|B) = \frac{(P(B|A) \times P(A))}{P(B)} \quad (1)$$

Keterangan:

B = Data dengan class yang belum diketahui.

A = Hipotesis Data B merupakan suatu class spesifik

P(A|B) = probabilitas hipotesis A berdasarkan kondisi B (posteriori prob.)

P(A) = Probabilitas hipotesis A (prior prob.)

P(B|A) = probabilitas B berdasarkan kondisi tersebut

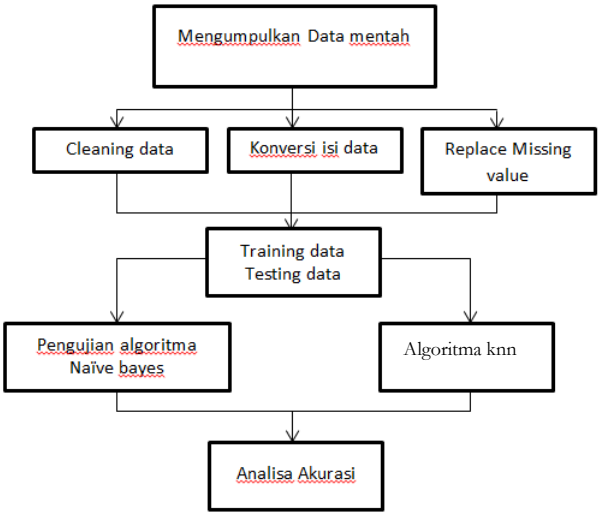
P(B) = probabilitas dari B

Sedangkan pada metode *K-Nearest Neighbors (KNN)* untuk mencari dekat atau jauhnya jarak antar titik pada kelas k=3 biasanya dihitung menggunakan jarak *Euclidean*. Jarak *Euclidean* adalah formula untuk mencari jarak antara 2 titik dalam ruang dua dimensi. Berikut rumus untuk menghitung jarak Euclidean:

$$d_{Euclidian}(x,) = \sqrt{\Sigma(xi - i)^2} \quad (2)$$

## Hasil dan Pembahasan

Pada tahap ini peneliti melakukan percobaaan dengan menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbors* (KNN) menggunakan bahasa pemograman *python* pada *google colab*. Berikut adalah tahap pengujian pada metode tersebut:




Gambar 1 Tahap Penelitian

- Pada tahap menggumpulkan data mentah peneliti mengambil data dari *website Universitas California Invene* (UCI) dengan jumlah data terkumpul sebanyak 3711 pasien terkena penyakit.
- *Cleaning data*, peneliti melakukan pengelolah data tiroid, mulai dari pengurangan data sehingga data yang digunakan peneliti hanya sebanyak 500 data pasien membuang duplikasi data dan menangani data yang hilang sehingga peneliti memperbaiki kesalahan pada data tersebut. Atribut yang digunakan pada data ini adalah *Age, sex, onthyroxine, query onthyroxine, on antithyroid medication, sick, pregnant, thyroid surgery, I131 treatment, query hypothyroid, query hyperthyroid, lithium, goiter, tumor, hypopituitary, psych, TSH, T3, TT4, T4U, FTI.*
- *Training data*, pada tahap ini peneliti mengabungkan data ke dalam format yang sesuai untuk proses pengujian

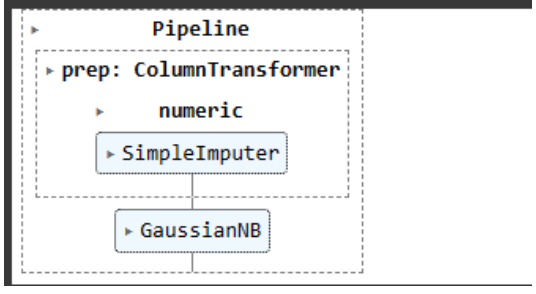
- Pengujian, setelah melakukan pengabungan data maka peneliti dapat menguji tingkat akurasi untuk melihat kinerja dari kedua algoritma tersebut. Pengujian tingkat akurasi menggunakan program *python* pada *google colab*.

```
<bound method NDFrame.head of       age  on thyroxine  query on thyroxine  on antithyroid medication  sick  \
0    41.0            0                   0                          0     0
1    23.0            0                   0                          0     0
2    46.0            0                   0                          0     0
3    70.0            1                   0                          0     0
4    70.0            0                   0                          0     0
..    ...          ...                 ...                        ...   ...
495  46.0            0                   0                          0     0
496  45.0            0                   0                          0     0
497  82.0            0                   1                          0     0
498  55.0            0                   0                          0     0
499  43.0            0                   0                          0     0

     pregnant  thyroid surgery  I131 treatment  query hypothyroid  lithium  \
0           0                0               0                  0        0
1           0                0               0                  0        0
2           0                0               0                  0        0
3           0                0               0                  0        0
4           0                0               0                  0        0
..        ...              ...             ...                ...      ...
495         0                0               0                  0        0
496         0                0               0                  0        0
497         0                0               0                  0        0
498         0                0               0                  0        0
499         0                0               0                  0        0

     goitre  tumor  hypopituitary  psych   TSH   T3    TT4   T4U    FTI
0         0      0              0      0  1.30  2.5  125.0  1.14  109.0
1         0      0              0      0  4.10  2.0  102.0  0.93  132.0
2         0      0              0      0  0.98  2.5  109.0  0.91  120.0
3         0      0              0      0  0.16  1.9  175.0  0.93  132.0
```

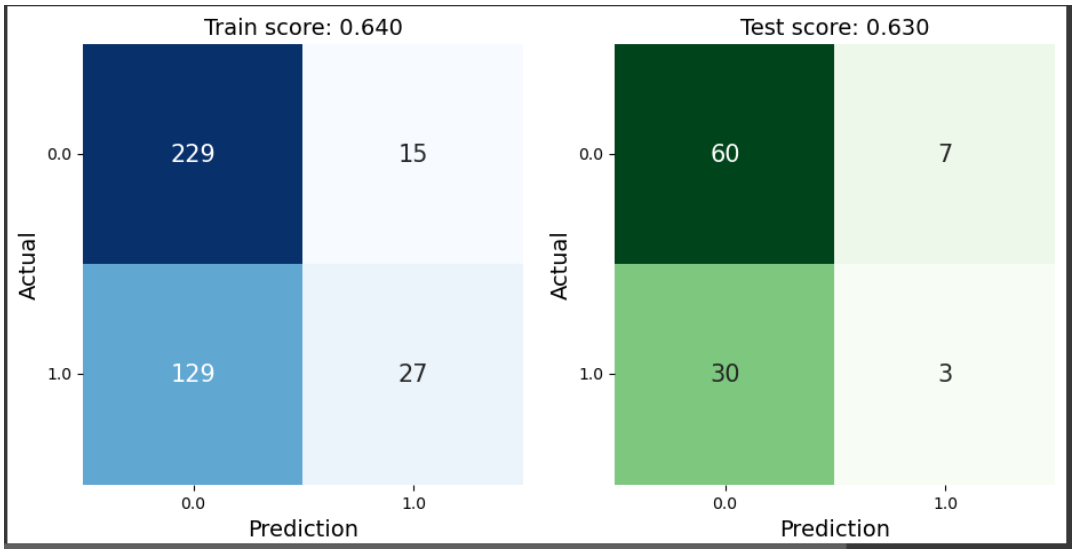
Gambar 2 Contoh Data

Pada tahap ini dilakukan pengujian dengan menggunakan dataset penyakit tiroid sebanyak 500 data dengan menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbors* (KNN). Pada tahap pengujian ini melakukan data training sehingga akan membahas algoritma yang diujikan, kemudian peneliti akan menganalisa dan dikomparasi. Algoritma dilakukan menggunakan pemograman *python* yang merupakan salah satu *freamwork* yang memiliki *library* yang luas dan didesain berorientasi objek. Pada gambar 2 menunjukan proses penerapan algoritma *Naïve Bayes* pada program *python*.

```
from sklearn.naive_bayes import GaussianNB
pipeline = Pipeline([
    ('prep', preprocessor),
    ('algo', GaussianNB())
])

pipeline.fit(X_train, y_train)
```




Gambar 3 Penerapan Algoritma *Naïve Bayes*




Gambar 4 Hasil *Accuracy Naïve Bayes*

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaler.fit(x_train)
x_train = scaler.transform(x_train)
x_test = scaler.transform(x_test)

#Library Pada Klasifikasi
from sklearn.neighbors import KNeighborsClassifier

#KNeighborsClassifier
modelKNN = KNeighborsClassifier(n_neighbors=4)
modelKNN.fit(x_train, y_train)

C:\Users\User\anaconda3\lib\site-packages\sklearn\nei
  return self._fit(X, y)
KNeighborsClassifier
KNeighborsClassifier(n_neighbors=4)
```

Gambar 5 Penerapan Algoritma *K-Nearest Neighbors* (KNN)

```
print(classification_report(y_test, y_pred))
```

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0.0 | 0.68 | 0.85 | 0.75 | 67 |
| 1.0 | 0.38 | 0.18 | 0.24 | 33 |
| accuracy | | | 0.63 | 100 |
| macro avg | 0.53 | 0.52 | 0.50 | 100 |
| weighted avg | 0.58 | 0.63 | 0.59 | 100 |

Gambar 6 Hasil *Accuracy K-Nearest Neighbors* (KNN)

## Kesimpulan

Pada kesimpulan yang telah dilakukan untuk Perbandingan Kinerja Algoritma *Naïve Bayes* Dan *K-Nearest Neighbors* (KNN) Dalam Memprediksi Penyakit Tiroid menghasilkan akurasi algoritma yang paling baik dalam menentukan identifikasi penyakit. Untuk mengukur kinerja kedua algoritma tersebut menggunakan data traning dan data testing sehingga menghasil kan akurasi sehingga perbandingan pada akurasi setiap algoritma *Naïve Bayes* dengan data testing 63% dan data training 64% sedangkan akurasi yang di hasilkan pada algoritma *K-Nearest Neighbors* (KNN) adalah 63% dapat disimpulkan bahwa algoritma yang memiliki klasifikasi sangat baik secara adalah *Naïve Bayes* dan *K-Nearest Neighbors* (KNN) berdasarkan penilaian akurasi dengan 500 data penyakit tiroid. Dengan begitu algoritma *Naïve Bayes* memberi pemecahan untuk permasalahan dalam mengidentifikasi penyakit tiroid.

## Daftar Rujukan

[1] Y. Timur, F. Simbolon, F. T. Informasi, and U. A. Indonesia, "Perancangan Aplikasi Untuk Memprediksi Seseorang Menderita Penyakit Hipertensi Menggunakan Data Mining Design of Application to Predict Someone Suffering Hypertension Using Data Mining," vol. 7, pp. 67–85, 2017.

[2] A. M. Widodo, Y. S. Anggraeni, N. Anwar, and ..., "Performansi K-NN, J48, Naive Bayes dan Regresi Logistik sebagai Algoritma Pengklasifikasi Diabetes," *Pros.* …, vol. 2007, 2021, [Online]. Available: http://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/253%0Ahttp://www.seminar.iaii.or.id/index.php/SISFOTEK/article/download/253/223.

[3] R. R. Al-hakim *et al.*, "SISTEM PAKAR UNTUK DIAGNOSIS PENYAKIT TIROID DENGAN GEJALA AN EXPERT SYSTEM FOR THYROID DISEASE DIAGNOSIS WITH PSYCHOLOGICAL SYMPTOMS DAN IT ' S ETHNOBOTANY TREATMENT," vol. 9, no. 7, 2022, doi: 10.25126/jtiik.202296763.

[4] Ayu Novita Sari, Natalia Silalahi, and Guidio Leonarde Ginting, "Sistem Pakar Diagnosa Penyakit Kelenjar Tiroid," *J. Ris. Komput.*, vol. 3, no. 2, pp. 18–20, 2018.

[5] A. Syahputri, A. Fauzi, and L. Arliana, "Implementasi Metode Certainty Factor Dalam Mendiagnosa Penyakit Tiroid," *J. Tek. Inform. Kaputama*, vol. 6, no. 1, pp. 306–318, 2022.

[6] M. Windarti, "Perbandingan Kinerja Algoritma Naïve Bayes Dan Bayesian Network Dalam Klasifikasi Masa Studi Mahasiswa," *Pros. Semin. Nas. Apl. Sains Teknol.*, no. September, pp. 249–261, 2018.

[7] M. Metode, F. Multiple, C. Decision, M. Fmcdm, and D. Yogyakarta, "Indonesian Journal of Business Intelligence," vol. 3, no. 2, pp. 54–60, 2020.

[8] L. Safitri, K. Cahayani, S. Chodidjah, and D. Indayanti, "Universitas Gunadarma, Jawa Barat, Indonesia," vol. 13, no. November, pp. 489–495, 2022.

[9] A. Supriyatna and W. P. Mustika, "Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, p. 152, 2018, doi: 10.30645/j-sakti.v2i2.78.

[10] B. Wijonarko, "Perbandingan Algoritma Data Mining Naive Bayes Dan Bayes Network Untuk Mengidentifikasi Penyakit Tiroid," *PILAR Nusa Mandiri*, vol. 14, no. 1, pp. 21–26, 2018, [Online]. Available: http://www.bsi.ac.id.

[11] E. S. R. Br.Situmorang, M. K. Anam, R. Rahmaddeni, and A. N. Ulfah, "Perbandingan Algoritma Svm Dan Nbc Dalam Analisa Sentimen Pilkada Pada Twitter," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 13, no. 3, p. 169, 2021, doi: 10.22303/csrid.13.3.2021.169-179.

[12] J. S. Informasi, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan

Metode Decision Tree," *J. IPTEK*, vol. 16, no. 1, pp. 1–7, 2017, [Online]. Available: http://jurnal.itats.ac.id/wp-content/uploads/2013/06/3.-BUDANIS-FINAL-hal-17-23.pdf.

[13] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.

[14] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2019.

[15] Aswan Supriyadi Sunge, "Prediksi Kompetensi Karyawan Menggunakan Algoritma C4 . 5 ( Studi Kasus : PT Hankook Tire Indonesia )," *Semin. Nas. Teknol. Inf. dan Komun. 2018 (SENTIKA 2018)*, vol. 2018, no. Sentika, pp. 23–24, 2018.

[16] E. YULIANTI and Y. A. NURDIN, "SISTEM PENDUKUNG KEPUTUSAN PENERIMAAN BANTUAN SISWA MISKIN (BSM) BERBASIS ONLINE DENGAN METODE KNN (K-NEAREST NEIGHBOR) (Studi kasus : SMPN 1 Koto XI Tarusan)," *J. Teknoif*, vol. 6, no. 1, pp. 12–17, 2018, doi: 10.21063/jtif.2018.v6.1.12-17.

[17] M. Reza Noviansyah, T. Rismawan, D. Marisa Midyanti, J. Sistem Komputer, and F. H. MIPA Universitas Tanjungpura Jl Hadari Nawawi, "Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya)," *J. Coding, Sist. Komput. Untan*, vol. 06, no. 2, pp. 48–56, 2018.