



## Komparasi Metode Naïve Bayes dan Kolmogorov Arnold Networks dalam Klasifikasi Tingkat Kemungkinan Obesitas

Renaya Aviary Azahra

*Fakultas Sains dan Teknologi, Universitas Darussalam Gontor*  
[renayaazahra42053@mhs.unida.gontor.ac.id](mailto:renayaazahra42053@mhs.unida.gontor.ac.id)

Triana Harmini

*Fakultas Sains dan Teknologi, Universitas Darussalam Gontor*  
 [triana@unida.gontor.ac.id](mailto: triana@unida.gontor.ac.id)

### ABSTRAK

Obesitas merupakan salah satu permasalahan pada bidang kesehatan yang cukup signifikan di seluruh dunia, dengan peningkatan prevalensi yang terus terjadi setiap tahunnya. Memprediksi kemungkinan obesitas dengan mendeteksi sedini mungkin merupakan Langkah yang sangat penting untuk mengurangi risiko komplikasi kesehatan yang terkait. Dengan adanya penelitian ini, bertujuan untuk mengkaji performa dua metode klasifikasi, yaitu Naïve Bayes dan Kolmogorov Arnold Networks (KAN). Analisis yang dilakukan dengan membandingkan akurasi dari masing-masing metode. Hasil menunjukkan bahwa metode Naïve Bayes lebih unggul dari Kolmogorov Arnold Networks. Hal ini mengindikasikan bahwa Naïve Bayes lebih efektif dalam data yang diteliti. Sebaliknya untuk Kolmogorov Arnold Networks menunjukkan keterbatasan pada dataset dengan fitur yang kurang kompleks.

Kata Kunci: Naïve Bayes, Kolmogorov Arnold Networks, Obesitas

### ABSTRACT

Obesity is one of the most significant health problems worldwide, with an increasing prevalence that continues to occur every year. Predicting the possibility of obesity by detecting it as early as possible is a very important step to reduce the risk of related health complications. With this study, the aim is to examine the performance of two classification methods, namely Naïve Bayes and Kolmogorov Arnold Networks (KAN). The analysis was carried out by comparing the accuracy of each method. The results show that the Naïve Bayes method is superior to Kolmogorov Arnold Networks. This indicates that Naïve Bayes is more effective in the data studied. Conversely, Kolmogorov Arnold Networks show limitations in datasets with less complex features.

Keywords: *Naïve Bayes, Kolmogorov Arnold Networks, Obesity*

**FAKULTAS TEKNIK – UNIVERSITAS ISLAM NUSANTARA**

Jalan Soekarno Hatta No. 530 Kota Bandung

## PENDAHULUAN

Permasalahan yang paling muncul dalam masyarakat yaitu masalah kesehatan. Salah satu masalah kesehatan yang cukup sering terjadi yaitu gangguan status gizi. Obesitas atau kelebihan berat badan merupakan penumpukan jaringan lemak yang abnormal atau berlebihan yang dapat memberi efek buruk pada kesehatan. Obesitas sering kali menjadi salah satu permasalahan gizi yang sering dihadapi oleh masyarakat global. Obesitas terjadi karena karena tingginya asupan makanan dengan kandungan gula dan lemak yang cukup tinggi. Apabila hasil dari makanan tersebut tidak dibakar atau dikeluarkan maka yang terjadi adalah penimbunan yang dapat menimbulkan efek berupa toksik bagi tubuh dan juga penyakit yang cukup beresiko[1].

Dengan ini, obesitas dapat terjadi apabila pola makan seseorang yang tidak seimbang atau keseringan mengkonsumsi makanan yang tidak sehat seperti *junk food*, makanan instant dan makanan tinggi kalori[2]. Obesitas atau tidaknya seseorang dapat dilihat dari berat badan dan juga tinggi badan yang kemudian diambil rata-ratanya menjadi Indeks Masa Tubuh. Dengan melihat data yang tersaji pada platform kaggle, masih banyak orang yang menderita obesitas, mulai dari usia muda hingga usia senja[3].

Salah satu metode klasifikasi pada bidang data mining yaitu Naïve Bayes, Naïve Bayes merupakan metode klasifikasi berdasarkan teorema popular bayes, yang biasa dikenal untuk membuat model sederhana khususnya di bidang klasifikasi dokumen dan prediksi penyakit. Pada dasarnya, algoritma Naïve Bayes merupakan algoritma pembelajaran mesin yang cenderung lebih banyak disarankan dibandingkan algoritma yang lain, dikarenakan memiliki kecepatan yang sangat tinggi dengan hasil analisis yang sangatlah baik.[4]

Selain memakai metode Naïve bayes, penelitian ini menggunakan metode Kolmogorov Arnold Networks. Merujuk pada aplikasi teorema Kolmogorov Arnold untuk representasi dan pemrosesan data dalam berbagai tugas data mining. Kolmogorov Arnold Networks merupakan konsep dalam teori jaringan neural yang berasal dari teorema representasi fungsi yang dikenal dengan teorema Kolmogorov Arnold. Teorema ini berkaitan dengan kemampuan jaringan neural untuk mewakili fungsi kontinu secara lebih efisien.[5]

## METODE PENELITIAN

Jenis penelitian ini menggunakan rancangan penelitian CRISP-DM (*Cross Industry Standart Process for Data Mining*). Metode ini memiliki 6 tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Proses-proses tersebut dijelaskan sebagai berikut :

a. *Business Understanding* (Pemahaman Bisnis)

Pada tahapan ini, pemahaman masalah yang terjadi dimasyarakat terkait pengaruh yang terjadi akibat obesitas. Selain memahami pengaruh, pemahaman mengapa obesitas itu dapat terjadi juga dilakukan. Dengan ini peneliti akan lebih memahami terkait faktor dan juga pengaruh yang akan terjadi akibat obesitas

b. *Data Understanding* (Pemahaman Data)

Tahapan ini, peneliti akan memahami data yang sudah diperoleh dari *platform Kaggle*. Data yang berisikan hal-hal terkait obesitas seperti berta badan, tinggi badan, BMI atau IMT, dan umur. Selain itu data yang digunakan berisikan 110 data. Pada data yang tertera, sudah tercantumkan terkait pernyataan tingkat obesitas yang tertera pada Label, yaitu *underweight*, *overweight*, *normal* dan *obesity.n*

c. *Data Preparation* (Persiapan Data)

Tahapan ini merupakan tahapan penting dalam CRISP-DM karena pada Langkah ini data mentah akan dipersiapkan agar sesuai untuk analisis dan pelatihan model. Pada tahap ini, data akan dibersihkan untuk mengatasi nilai yang hilang, menghapus data duplikat juga mengoreksi kesalahan data. Pada tahap ini juga, data akan dibagi menjadi set pelatihat (*training set*) dan set pengujian (*test set*).

d. *Modelling* (Permodelan)

Tahapan ini bertujuan untuk membuat model prediktif berdasarkan data yang telah diproses. Proses ini melibatkan beberapa tahapan yang bertujuan untuk mengklasifikasi data baru, seperti memilih model berdasarkan jenis masalah. Pada penelitian ini akan digunakan model klasifikasi *Naïve Bayes* dan *Kolmogorov Arnold Networks*.

e. *Evaluation* (Pengujian)

Tahapan ini dilakukan untuk mengevaluasi dengan metrik seperti akurasi untuk klasifikasi. Pada tahap ini juga dilakukan peninjauan kembali seluruh proses dari tahap persiapan data hingga pelatihan model untuk memastikan tidak ada data yang terlewat atau dilakukan secara kurang benar

f. Deployment (Penyebaran)

Pada tahapan ini akan dibuat laporan serta visualisasi data agar lebih mudah dipahami.

## HASIL DAN DISKUSI

Tujuan dari penelitian ini adalah melakukan klasifikasi Tingkat obesitas dengan kategori *underweight*, *normal*, *overweight*, dan *obese*. Data yang sudah terklasifikasi tersebut, kemudian dibandingkan dengan nilai akurasi dengan model Naïve Bayes dan Kolmogorov Arnold Networks. Pada tahapan ini akan diuraikan pembahasan mengenai implementasi kedua metode tersebut untuk data obesitas ini.

### Hasil

Tujuan dari penelitian ini adalah melakukan klasifikasi Tingkat obesitas dengan kategori *Underweight*, *Normal*, *Overweight*, *Obesity*. Data yang sudah terklasifikasi kemudian dibandingkan nilai akurasi dengan metode *Naïve Bayes* dan *Kolmogorov Arnold Networks*. Pada tahapan ini akan diuraikan pembahasan mengenai hasil implementasi kedua metode tersebut pada dataset obesitas ini. Berikut uraian tahapan untuk hasil penelitian sebagai berikut :

1. Pengumpulan data

Data yang dipakai untuk penelitian ini didapatkan dari *platform* Kaggle.com dengan kata kunci *Obesity Classification Dataset* dengan total data 1000. Dengan label *Weight*, *Height*, *BMI*.

2. Pra-pemrosesan Data

- Data Cleaning : Pada tahapan ini, dilakukan penghapusan data yang tidak relevan, data duplikat, atau mengisi nilai hilang
- Data Transformation : Tahapan ini dilakukan perubahan terhadap format data agar sesuai untuk analisis misalnya dengan normalisasi

### 3. Pemilihan Fitur

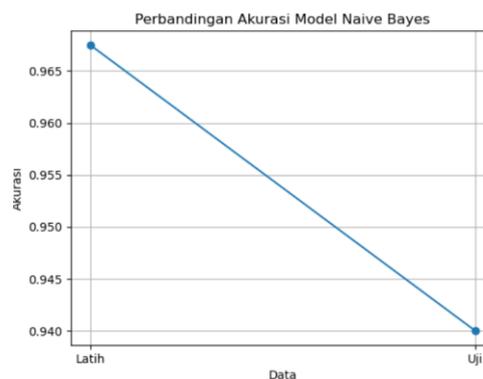
Tujuan dari pemilihan fitur atau variable guna memilih yang paling relevan untuk tujuan analisis. Dengan menyederhanakan data sehingga dapat meningkatkan efisiensi algoritma tanpa mengorbankan akurasi model.

### 4. Pembentukan Model

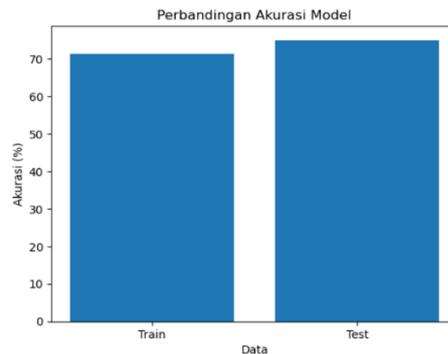
Pada tahapan ini dilakukan pemilihan dan Pembangunan model algoritma yang sesuai dengan tujuan penelitian ini. Model yang dipakai untuk penelitian ini yaitu klasifikasi

### 5. Evaluasi Model

Evaluasi model dilakukan guna mengukur performa yang digunakan dalam penelitian ini. Diukur menggunakan metrik yang sesuai, salah satunya adalah akurasi. Berikut hasil akurasi dari setiap algoritma



Gambar 1



Gambar 2

## 6. Analisis Hasil

Tahapan ini bertujuan untuk menganalisis dan menginterpretasikan hasil model berdasarkan tujuan penelitian. Pada tahapan ini, hasil dievaluasi untuk mengetahui seberapa baik model bekerja dalam menemukan pola atau memberikan prediksi

### Diskusi

Pada penelitian ini, metode klasifikasi yang digunakan yaitu Kolmogorov Arnold Networks. Dalam konteks klasifikasi kemungkinan obesitas dengan berdasarkan beberapa variable terkait. Dengan hasil menunjukkan bahwa Naïve Bayes berhasil mencapai akurasi klasifikasi sebesar 94% dan Kolmogorov Arnold Networks dengan hasil yang hanya 75%.

Naïve bayes yang memiliki performa yang lebih baik dapat dijelaskan melalui beberapa faktor. Dengan asumsi, setiap fitur dianggap independent. Meski asumsi ini tidak sepenuhnya benar dalam konteks data Kesehatan yang cukup kompleks, terbukti efektif dalam penelitian ini. Dengan kata lain, Naïve bayes mampu menangkap pola yang cukup relevan dalam data obesitas. Selain itu Naïve bayes memiliki keunggulan dalam menangani dataset dengan variable kategori.

Di sisi lain, Kolmogorov Arnold Networks yang menggunakan pendekatan pemetaan non-linear cenderung terbilang kurang optimal dalam menangani klasifikasi untuk dataset ini. Dengan akurasi sebesar 75% menunjukkan bahwa KAN mungkin kurang cocok dalam klasifikasi kemungkinan obesitas berdasarkan variable yang ada. Salah satu kemungkinan penyebabnya adalah Tingkat sensitivitas KAN terhadap parameter dan konfigurasi jaringan.

Kompleksitas jaringan dalam KAN yang berpotensi menyebabkan overfitting pada data yang relative sederhana yang pada akhirnya menurunkan akurasi

Dari hasil yang terlihat bahwa untuk penelitian klasifikasi Tingkat kemungkinan obesitas pada dataset tertentu, Naïve Bayes lebih unggul dibandingkan dengan Kolmogorov Arnold Networks. Hal ini cukup menyoroti betapa pentingnya pemilihan model yang sesuai dengan karakteristik data dan tujuan klasifikasi. Secara sederhana, model yang lebih sederhana namun efektif seperti Naïve Bayes contohnya dapat memberikan hasil yang lebih baik pada kasus dimana fitur cenderung bersifat kategorikal atau tidak terlalu banyak berinteraksi dalam hal kompleksitas.

## **KESIMPULAN**

Hasil komparasi antara metode Naïve Bayes dan Kolmogorov Arnold Networks (KAN) untuk klasifikasi Tingkat kemungkinan obesitas, dapat disimpulkan bahwa metode Naïve Bayes menunjukkan performa yang lebih baik dengan akurasi 94% dibandingkan KAN yang hanya mencapai 75%. Keunggulan pada Naïve Bayes ini terutama disebabkan oleh kemampuannya menangani dataset dengan variable kategori secara efektif dan sederhana.

Sementara itu, Kolmogorov Arnold Networks kurang optimal untuk penelitian ini. Diprediksi karena sensitivitas terhadap konfigurasi parameter serta risiko overfitting pada dataset yang tidak kompleks. Hasil ini menekankan pentingnya pemilihan model yang tepat berdasarkan karakteristik data. Untuk klasifikasi ini, metode sederhana seperti Naïve Bayes dapat memberikan performa yang cukup baik dibandingkan metode yang lebih kompleks

Penelitian selanjutnya dengan variasi dataset atau kombinasi metode lain disarankan untuk menguji konsistensi dan generalisasi dari hasil ini.

## DAFTAR PUSTAKA

- [1] S. Arifani and Z. Setiyaningrum, “Faktor Perilaku Berisiko yang Berhubungan Dengan Kejadian Obesitas Pada Usia Dewasa di Provinsi Banten Tahun 2018,” *J. Kesehat.*, vol. 14, no. 2, pp. 160–168, 2021, doi: 10.23917/jk.v14i2.13738.
- [2] Fernando Trisno Swandi, Restyandito, and K. Adi Nugraha, “Analisa Usability Desain Aplikasi Controlling Calories untuk Mendapatkan Berat Badan Ideal,” *Jnanaloka*, pp. 19–26, 2023, doi: 10.36802/jnanaloka.2023.v4-no01-19-26.
- [3] N. R. Hanani Retno, Badrah Siti, “Pola Makan , Aktivitas Fisik dan Genetik Mempengaruhi Kejadian Obesitas,” vol. 14, pp. 120–129, 2021.
- [4] A. Ridwan, “Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [5] Z. Liu *et al.*, “KAN: Kolmogorov-Arnold Networks,” pp. 1–48, 2024, [Online]. Available: <http://arxiv.org/abs/2404.19756>.