

# Analisis Prediksi Level Obesitas Menggunakan Perbandingan Algoritma *Machine Learning* dan *Deep Learning*

Lila Setiyani<sup>1#</sup>, Ayu Nur Indahsari<sup>2</sup>, Rusdianto Roestam<sup>3</sup>

<sup>1</sup>STMIK ROSMA

Jl. Parahiyangan Karawang, Indonesia

<sup>2,3</sup>President University

Jababeka Education Park Bekasi, Indonesia

#lila.setiyani@dosen.rosma.ac.id

---

---

## Abstrak

Level obesitas dapat diidentifikasi berdasarkan kebiasaan makan dan kondisi fisik yang terdiri dari beberapa parameter yang dapat digunakan untuk mengukur tingkat akurasi dari prediksi level obesitas. Penelitian ini bertujuan untuk membuat prediksi level obesitas dengan akurasi yang baik. Metode yang digunakan adalah dengan membandingkan algoritma machine learning dan deep learning, dataset diambil dari data level obesitas pada individu dari negara Mexico, Peru dan Kolombia yang didasarkan pada kebiasaan makan dan kondisi fisik, data tersebut terdiri dari 17 atribut dan 2111 record. Berdasarkan hasil analisis algoritma *machine learning* didapati akurasi dari *random forest* sebesar 96,37%, *decision tree classifier* sebesar 88,33%, *logistic regression* sebesar 73,66%, *naïve bayes* sebesar 56% dan KNN sebesar 88,96%. Sedangkan *deep learning* didapati akurasi sebesar 86,05%. Algoritma *machine learning random forest* memiliki akurasi yang paling baik dan dapat memprediksi presentase akurasi serta level obesitas.

**Kata kunci:** obesitas, *machine learning*, *deep learning*, algoritma, prediksi

## Abstract

*Obesity levels can be identified based on eating habits and physical conditions which consist of several parameters that can be used to measure the accuracy of the prediction of obesity levels. This study aims to predict obesity levels with good accuracy. The method used is to compare machine learning and deep learning algorithms, the dataset is taken from obesity level data in individuals from Mexico, Peru and Colombia based on eating habits and physical condition, the data consists of 17 attributes and 2111 records. Based on the results of machine learning algorithm analysis, it was found that the accuracy of random forest was 96.37%, decision tree classifier was 88.33%, logistic regression was 73.66%, naïve Bayes was 56% and KNN was 88.96%. Meanwhile, deep learning has an accuracy of 86.05%. The random forest machine learning algorithm has the best accuracy and can predict the percentage of accuracy and the level of obesity*

**Keywords:** obesity, *machine learning*, *deep learning*, algorithm, prediction

---

---

## I. PENDAHULUAN

Obesitas merupakan kondisi kelebihan berat badan yang membawa manifestasi negative bagi Kesehatan, kondisi ini ditentukan dari parameter Indeks Massa Tubuh (IMT) [1]. Puspitasari menjelaskan bahwa obesitas di pengaruhi oleh jenis kelamin, tingkat oengetahuan, tingkat pendidikan, jenis pekerjaan, status pernikahan, Riwayat keturunan, aktivitas fisik, status merokok dan intake

kalori [2]. Obesitas tidak hanya terjadi pada usia dewasa [3], dan remaja [4], namun juga terjadi di usia anak – anak [5]. Menurut Fabio, ada tujuh level dalam obesitas diantaranya *Insufficient Weight*, *Normal Weight*, *Overweight Level I*, *Overweight Level II*, *Obesity Type I*, *Obesity Type II*, dan *Obesity Type III* [6]. Akibat dari obesitas ini membawa beberapa penyakit diantaranya Sindroma Metabolic yang memicu penyakit Cardiovascular, diabetes [3], sedangkan pada anak berdampak pada

masalah perkembangan fisik motorik [7]. Melihat besarnya dampak negative yang di sebabkan oleh obesitas, ini menarik peneliti untuk menyusun algoritma yang dapat memprediksi level obesitas.

Untuk membuat prediksi, beberapa peneliti menggunakan machine learning dan deep learning. Machine learning mengacu secara luas pada proses penyesuaian model prediktif dengan data atau mengidentifikasi pengelompokan informatif dalam data [8]. Mesin learning juga telah membantu mengotomatisasi penelitian dengan menggunakan model yang dipelajari mesin untuk menyarankan eksperimen atau simulasi baru, machine learning ini menghasilkan loop umpan balik data dari pembuatan data, pelatihan ulang model dan prediksi yang ditingkatkan mewakili perubahan paradigma dalam bahan penelitian [9]. Sedangkan deep learning merupakan konsep dari machine learning yang didasarkan pada jaringan saraf tiruan, menurut Janiesch, untuk banyak aplikasi model *deep learning* mengungguli model *machine learning* [10]. Mahendra *et al* dalam penelitiannya menggunakan *machine learning* dalam prediksi ekspor tanaman obat, aromatic dan rempah – rempah [11]. Saiful *et al* dalam penelitiannya menggunakan *machine learning* dengan algoritma *linier regression* pada prediksi harga rumah [12]. Muhamad *et al* juga menggunakan algoritma machine learning dalam penentuan model prediksi produksi telur ayam petelur [13]. Berdasarkan penelitian diatas belum ada penelitian yang membandingkan tingkat akurasi antara algoritma machine learning dengan *deep learning*. Oleh sebab itu dalam prediksi level obesitas ini, peneliti akan memperlihatkan perbandingan akurasi antara dua algoritma tersebut, sehingga dapat dipilih algoritma terbaik dengan presentase akurasi terbesar yang dapat diterapkan.

## II. METODE PENELITIAN

Prosedur dalam penelitian ini adalah seperti pada Gambar 1. Prosedur pada penelitian diawali dengan mengambil dataset dari UCI, yaitu dataset

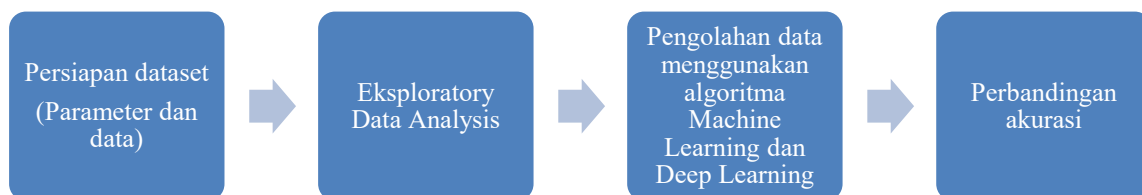
estimasi dari obesitas level pada individu dari negara Mexico, Peru dan Columbia [14]. Berdasarkan dataset tersebut didapati ada 17 atribut dengan 2111 *record*. Tools yang digunakan pada penelitian ini adalah [colab.research.google.com](http://colab.research.google.com) [15]. Dataset selanjutnya di import dan dilanjutkan dengan proses Exploratory Data Analysis (EDA) [16], kemudian memproses data menggunakan algoritma *machine learning* dengan tipe klasifikasi dan menguji akurasi dengan beberapa teknik yaitu *Random Forest Classifier* [17], *Decision Tree Classifier* [18], *Logistic Regression*, *Multinomial Naïve Bayes*, dan *KNN Classifier*. Selanjutnya dilakukan pemrosesan data dengan algoritma *deep learning*. Berdasarkan hasil pemrosesan *machine learning* dan *deep learning* kemudian dibandingkan nilai akurasinya. Model yang paling baik adalah model yang memiliki akurasi terbesar.

Parameter yang digunakan untuk prediksi obesitas pada penelitian ini adalah *eating habit (frequent consumption of high caloric food (FAVC), frequency of consumption of vegetables (FCVC), number of main meals (NCP), consumption of food between meals (CAEC), consumption of water daily (CH20), and consumption of alcohol (CALC)), physical condition (calories consumption monitoring (SCC), physical activity frequency (FAF)), time using technology devices (TUE), dan transportation used (MTRANS)* [6].

### A. Random Forest Classifier

*Random forest* adalah algoritma klasifikasi dan regresi yang merupakan pengembangan dari *decision tree*, dimana setiap *decision tree* tersebut dilakukan proses pelatihan menggunakan sampel individu [19]. Kriteria yang digunakan dalam *random forest* pada penelitian ini menggunakan Gini dan Entropy seperti pada persamaan (1) dan (2).

$$Gini = 1 - \sum_{i=1}^c (p_i)^2 \quad (1)$$



Gambar 1. Prosedur penelitian

Keterangan:  
 $p$  = probabilitas

$$E(S) = -p(+)\log p(+)-p(-)\log p(-) \quad (2)$$

Keterangan:  
 $E$  = Entropy  
 $p$  = probabilitas

### B. Decision Tree Classifier

*Decision tree* digunakan untuk mempelajari klasifikasi dan prediksi pola dari data dan menggambarkan relasi dari variabel atribut  $x$  dan variabel dalam bentuk pohon [20].

### C. Logistic Regression

Regresi logistik digunakan untuk mengidentifikasi apakah terdapat hubungan antara variabel dependen yang terkait dengan kejadian, ya atau tidak dari suatu peristiwa dengan satu atau lebih variabel independen yang bertipe *categorical* atau *continuous*. Regresi logistik secara matematis memperkirakan fungsi regresi linier berganda [21].

### D. Naïve Bayes Classifier

Naïve Bayes merupakan pengklasifikasian statistik yang digunakan untuk memprediksi probabilitas keanggotaan suatu class yang didasarkan pada teorema Bayes yang memiliki bentuk umum sebagai berikut [19]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (3)$$

dimana:

$X$  = data dengan class yang belum diketahui

$H$  = Hipotesis data  $X$  merupakan satu class spesifik

$P(H|X)$  = probabilitas hipotesis  $H$  berdasarkan kondisi  $X$

$P(H)$  = Probabilitas hipotesis  $H$

$P(X|H)$  = Probabilitas hipotesis  $X$  berdasarkan kondisi pada hipotesis  $H$

$P(X)$  = probabilitas dari  $X$

Keuntungan dari penggunaan Naïve Bayes adalah metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk

menentukan estimasi parameter yang diperlukan dalam proses klasifikasi [22].

### E. KNN – Classifier

KNN merupakan proses untuk mengelompokkan data ke dalam kelas-kelas yang telah ditentukan sebelumnya berdasarkan jarak terdekat/tingkat kemiripan data tersebut dengan dataset/data latih yang ada [22]. KNN termasuk pada kelompok *instance-based learning* yang merupakan salah satu teknik *lazy learning* yang dilakukan dengan mencari kelompok  $k$  objek dalam *data training* yang paling dekat atau mirip dengan objek pada data baru atau *data testing* [23].

### F. Deep Learning

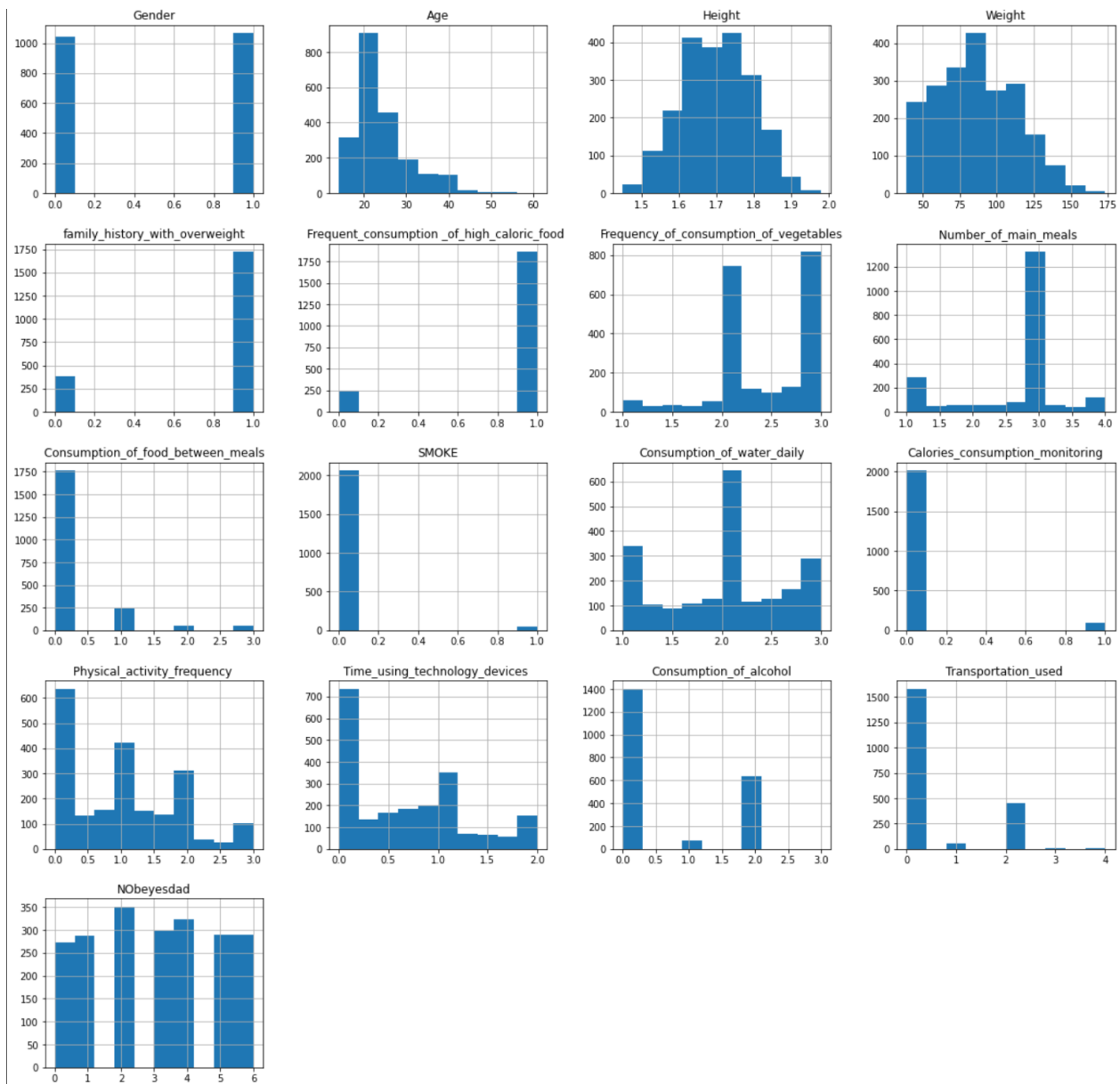
*Deep learning* merupakan cabang dari *machine learning* berbasis jaringan saraf, dimana sebuah komputer belajar mengklasifikasikan secara langsung dari gambar atau suara. Pada penggunaan teknik *deep learning* digunakan model dengan 2 dua layer, semakin banyak layer yang digunakan akan mempengaruhi dari tingkat akurasi yang dihasilkan.

## III. HASIL DAN PEMBAHASAN

Berdasarkan data yang di ambil dari UCI beberapa parameter belum memiliki *value* dalam *numeric*, maka pada proses awal setelah data diimport ke dalam [colab.research.google.com](https://colab.research.google.com) dilakukan perubahan *value data* pada setiap atribut atau parameter diantaranya *atribut gender*, *family\_history\_with\_overweight*, *frequent\_consumption\_of\_high\_caloric\_food*, *consumption\_of\_food\_between\_meals*, *SMOKE*, *calories\_consumption\_monitoring*, *consumption\_of\_alcohol*, *transportation\_used* dan *NObedesdad*

### A. Exploratory Data Analysis

Untuk mengetahui gambaran data pada *exploratory data analysis*, data dari setiap atribut di visualisasikan ke dalam bentuk grafik.



**Gambar 2. Graphic visualisasi atribut**

Gambar 2 memperlihatkan *value female* lebih banyak dibandingkan *male*, umur terbanyak pada data adalah kisaran 20-30 tahun, *height* terbanyak adalah antara 1,6 – 1,8, kebanyakan data memiliki *weight* 50 - 125, dengan nilai tertinggi mendekati 175 dan kebanyakan memiliki keluarga dengan riwayat *overweight*, serta kebanyakan mengkonsumsi kalori tinggi. Pada konsumsi *vegetable* memperlihatkan *value* “always” hampir sama dengan *value* “no” dan jumlah *main meals* sebanyak 3. Pada *consumption of food between meals* memperlihatkan *value* “sometimes” terbanyak dibandingkan *value* lainnya. Sedangkan terkait atribut meroko, didapati hampir semua merokok. Pada *consumption of water daily* memperlihatkan *value* tertinggi adalah 2, sedangkan pada *calories consumption monitoring*

memplihatkan hampir semua *value* menyatakan “no” yang berarti tidak memonitor konsumsi kalori. Pada *physical activity frequency* didapati *value* tertinggi adalah 0, ini menggambarkan tidak adanya *physical activity*. Pada *time using technology device* terbanyak memiliki *value* “0”. Pada *consumption of alcohol* memperlihatkan *value* “sometime” tertinggi. Pada *transportation used* *value public transportation* tertinggi, dan grafik *NObesydad* memperlihatkan level dari obesitas yang didapat dari dataset.

Berdasarkan hasil pengolahan data, didapati korelasi antar atribut seperti pada Tabel 1.

Tabel 1. Korelasi antar atribut parameter

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
A	1,0	0,0	0,6	0,2	0,1	0,1	-0,3	0,1	-0,0	0,0	0,1	-0,1	0,2	0,0	0,0	0,2	0,0
B	0,0	1,0	-0,0	0,2	0,2	0,1	0,0	-0,0	-0,1	0,1	-0,0	-0,1	-0,1	-0,3	-0,0	0,6	0,2
C	0,6	-0,0	1,0	0,5	0,2	0,2	-0,0	0,2	-0,1	0,1	0,2	-0,1	0,3	0,1	-0,1	0,1	0,0
D	0,2	0,2	0,5	1,0	0,5	0,3	0,2	0,1	-0,3	0,0	0,2	-0,2	-0,1	-0,1	-0,3	-0,0	0,4
E	0,1	0,2	0,2	0,5	1,0	0,2	0,0	0,1	-0,3	0,0	0,1	-0,2	-0,1	0,0	0,0	0,1	0,3
F	0,1	0,1	0,2	0,3	0,2	1,0	-0,0	-0,0	-0,2	-0,1	0,0	-0,2	-0,1	0,1	-0,1	-0,0	0,0
G	-0,3	0,0	-0,0	0,2	0,0	-0,0	1,0	0,0	-0,0	0,0	0,1	0,1	0,0	-0,1	-0,1	-0,1	0,0
H	0,1	-0,0	0,2	0,1	0,1	-0,0	0,0	1,0	-0,0	0,0	0,1	-0,0	0,1	0,0	-0,1	0,1	-0,1
I	-0,0	-0,1	-0,1	-0,3	-0,3	-0,2	-0,0	-0,0	1,0	0,1	0,0	0,2	0,0	-0,0	0,1	-0,0	-0,2
J	0,0	0,1	0,1	0,0	0,0	-0,1	0,0	0,0	0,1	1,0	-0,0	0,0	0,0	0,0	-0,0	0,0	-0,0
K	0,1	-0,0	0,2	0,2	0,1	0,0	0,1	0,1	0,0	-0,0	1,0	0,0	0,2	0,0	-0,1	-0,0	0,1
L	-0,1	-0,1	-0,1	-0,2	-0,2	-0,2	0,1	-0,0	0,2	0,0	0,0	1,0	0,1	-0,0	0,0	-0,0	-0,1
M	0,2	-0,1	0,3	-0,1	-0,1	-0,1	0,0	0,1	0,0	0,0	0,2	0,1	1,0	0,1	0,1	0,0	-0,1
N	0,0	-0,3	0,1	-0,1	0,0	0,1	-0,1	0,0	-0,0	0,0	0,0	-0,0	0,1	1,0	0,1	-0,2	-0,1
O	0,0	-0,0	-0,1	-0,3	0,0	-0,1	-0,1	-0,1	0,1	-0,0	-0,1	0,0	0,1	0,1	1,0	0,1	-0,1
P	0,2	0,6	0,1	-0,0	0,1	-0,0	-0,1	0,1	-0,0	0,0	-0,0	-0,0	0,0	-0,2	0,1	1,0	0,0
Q	0,0	0,2	0,0	0,4	0,3	0,0	0,0	-0,1	-0,2	-0,0	0,1	-0,1	-0,1	-0,1	-0,1	0,0	1,0

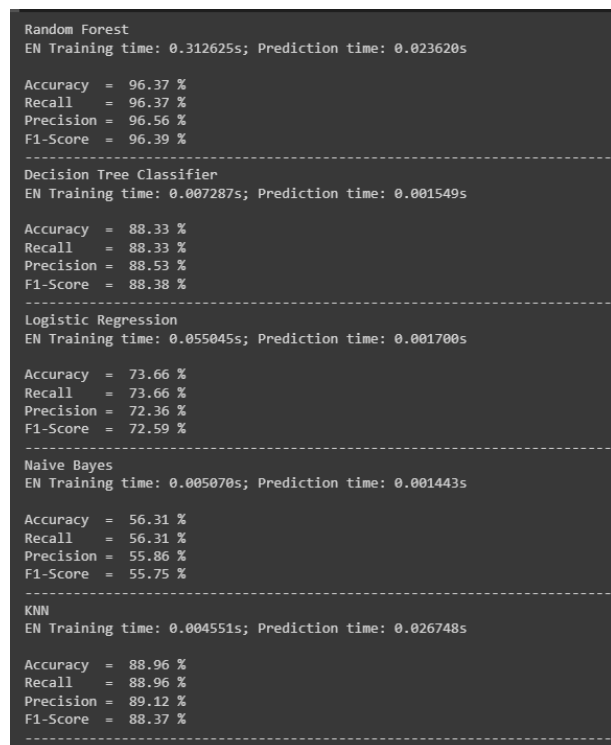
Keterangan:

- A. Gender
- B. Age
- C. Height
- D. Weight
- E. Family\_history\_with\_overweight
- F. Frequent\_consumption\_of\_high\_caloric\_food
- G. Frequent\_of\_consumption\_of\_vegetables
- H. Number\_of\_main\_meals
- I. Consumption\_of\_food\_between\_meals
- J. SMOKE
- K. Consumption\_of\_water\_daily
- L. Calories\_consumption\_monitoring
- M. Physical\_activity\_frequency
- N. Time\_using\_technology
- O. Consumption\_of\_alcohol
- P. Transportation\_used
- Q. Nobeyesdad

Tabel 1 dapat dimaknai bahwa parameter yang paling mempengaruhi NObesyesdad adalah *age*, *weight*, dan *family\_historic\_overweight*, karena memiliki nilai korelasi  $\geq 0$ .

**B. Classification (Machine Learning)**

Berdasarkan hasil pengujian menggunakan *random forest classifier* didapati akurasi skor sebesar 0,96 atau 96%. Berdasarkan hasil dari pemrosesan data menggunakan teknik *decision tree classifier* didapati bahwa akurasi skor sebesar 0,95 atau 95%. Berdasarkan hasil pemrosesan data menggunakan teknik *logistic regression* didapati akurasi skor sebesar 0,83 atau 83%. Berdasarkan hasil pemrosesan data menggunakan teknik *naive bayes* didapati akurasi skor sebesar 0,56 atau 56%

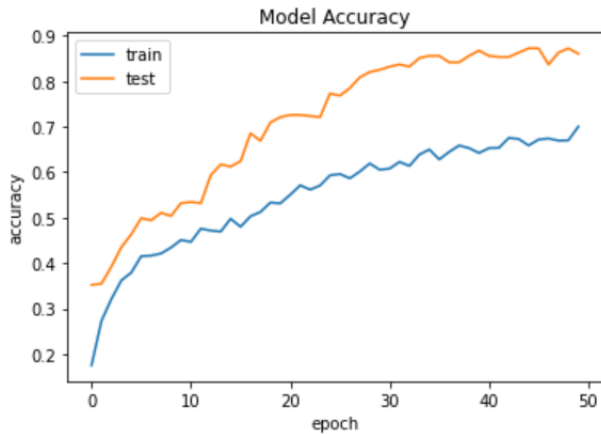


Gambar 3. Perbandingan akurasi algoritma machine learning

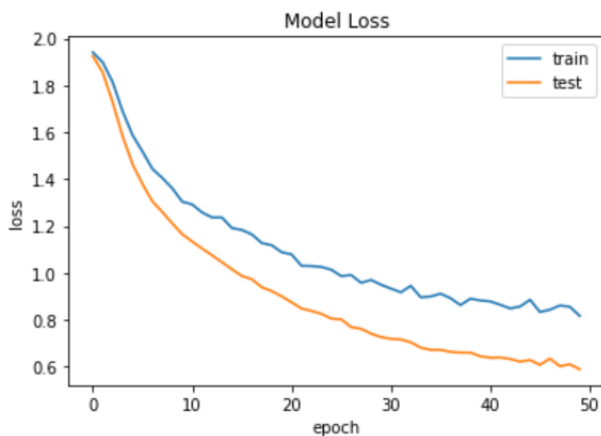
Berdasarkan hasil pemrosesan data menggunakan teknik KNN Classifier didapati akurasi skor sebesar 0,92 atau 92%. Berdasarkan 5 (lima) teknik *machine learning* yang telah digunakan, selanjutnya dapat dilihat perbandingan skor akurasi pada setiap teknik seperti pada Gambar 3. Berdasarkan hasil yang ada dapat terlihat bahwa teknik *random forest classifier* memiliki akurasi yang paling baik.

### C. Pengujian Deep Learning

Gambar 4 dan Gambar 5 menunjukkan evaluasi akurasi model dan *lost model* dari algoritma *deep learning*. Berdasarkan hasil dari grafik *model loss* dan *model accuracy* dapat dimaknai bahwa model memberikan hasil yang baik hal ini dapat dilihat dari garis antara *train* dan *test* yang hampir berhimpitan. Gambar 6 menunjukkan hasil akurasi menggunakan teknik *deep learning*.



Gambar 4. Hasil evaluasi *model accuracy*



Gambar 5. Hasil evaluasi *loss model*

```
[ ] # Evaluasi Model

from sklearn.metrics import classification_report, accuracy_score

categorical_pred = np.argmax(model.predict(x_test), axis=1)

print('Results for Categorical Model')
print(accuracy_score(y_test_eval, categorical_pred))
print(classification_report(y_test_eval, categorical_pred))

14/14 [=====] - 0s 2ms/step
Results for Categorical Model
0.8685208945626478
      precision    recall  f1-score   support

 0.0         0.77     0.98     0.86         54
 1.0         0.88     0.45     0.62         58
 2.0         0.87     0.94     0.90         70
 3.0         0.95     0.98     0.97         60
 4.0         0.98     0.98     0.98         65
 5.0         0.72     0.88     0.79         58
 6.0         0.90     0.74     0.81         58

 accuracy          0.86         423
 macro avg         0.87         423
 weighted avg      0.87         423
```

Gambar 6. Hasil akurasi algoritma *deep learning*

Tabel 2. Evaluasi model

Teknik	Skor Akurasi	% Akurasi
<i>Random Forest Classifier</i>	0,96	96%.
<i>Decision Tree Classifier</i>	0,95	95%
<i>Logistic Regression</i>	0,83	83%
<i>Multinomial Naïve Bayes</i>	0,56	56%
<i>KNN Classifier</i>	0,92	92%
<i>Deep Learning</i>	0,86	86%

Berdasarkan hasil evaluasi dari metode algoritma *machine learning* dan *deep learning* didapati perbandingan akurasi seperti ditunjukkan pada Tabel 2. Berdasarkan hasil evaluasi model, didapati bahwa *random forest classifier* adalah teknik yang dapat dipilih sebagai model prediksi level obesitas, karena memiliki skor akurasi tertinggi dibandingkan teknik lain yang telah diujikan.

### IV. KESIMPULAN

Algoritma *machine learning* maupun *deep learning* terbukti dapat digunakan untuk membuat prediksi dari level obesitas. Pada setiap teknik yang digunakan dari algoritma tersebut, didapati memiliki skor akurasi yang berbeda-beda. *Random forest classifier* memiliki skor akurasi tertinggi, sehingga model tersebut dapat diterapkan untuk dapat memprediksi level obesitas. Penelitian ini, dapat membuka wawasan untuk penerapan di bidang lain yang membutuhkan analisis prediksi, klasifikasi serta estimasi.

### REFERENSI

- [1] D. A. Christianto, “Hubungan Aktivitas Fisik Terhadap Kejadian Obesitas Berdasarkan Indeks Massa Tubuh Di Desa Banjaroyo,” *Berk. Ilm. Kedokt. Duta Wacana*, vol. 3, no. 2, p. 78, 2018.
- [2] N. Puspitasari, “Faktor kejadian obesitas sentral paa usia dewasa,” *Higeia J. puplic Heal. Res. Dev.*, vol. 2, no. 2, pp. 249–259, 2018.
- [3] R. V. Suryadinata and D. A. Sukarno, “Pengaruh Aktivitas Fisik Terhadap Risiko Obesitas Pada Usia Dewasa,” *Indones. J. Public Heal.*, vol. 14, no. 1, pp. 106–116, 2019.
- [4] A. Mutia, J. Jumiyati, and K. Kusdalinah, “Pola Makan Dan Aktivitas Fisik Terhadap Kejadian Obesitas Remaja Pada Masa Pandemi Covid-19,” *J. Nutr. Coll.*, vol. 11, no. 1, pp. 26–34, 2022.
- [5] W. Damopoli, N. Mayulu, and G. Masi, “Hubungan Konsumsi Fastfood Dengan Kejadian Obesitas Pada Anak Sd Di Kota Manado,” *Ejournal Keperawatan*, vol. 1, no. 1, pp. 1–7, 2013.
- [6] F. M. Palechor and A. de la H. Manotas, “Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from

- Colombia, Peru and Mexico,” *Data Br.*, vol. 25, p. 104344, 2019.
- [7] L. N. W. F. Fajzrina and R. R. Diana, “ANALISIS DAMPAK OBESITAS TERHADAP PERKEMBANGAN FISIK MOTORIK ANAK USIA 5 TAHUN Lati,” *J. Progr. Stud. Pendidik. Anak Usia Dini “CERIA,”* vol. 11, no. 1, pp. 62–74, 2022.
- [8] J. G. Greener, S. M. Kandathil, L. Moffat, and D. T. Jones, “A guide to machine learning for biologists,” *Nat. Rev. Mol. Cell Biol.*, vol. 23, no. 1, pp. 40–55, 2022.
- [9] G. L. W. Hart, T. Mueller, C. Toher, and S. Curtarolo, “Machine learning for alloys,” *Nat. Rev. Mater.*, vol. 6, no. 8, pp. 730–755, 2021.
- [10] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electron. Mark.*, vol. 31, no. 3, pp. 685–695, 2021.
- [11] M. Mahendra, R. Chandra Telaumbanua, A. Wanto, and A. Perdana Windarto, “Akurasi Prediksi Ekspor Tanaman Obat, Aromatik dan Rempah-Rempah Menggunakan Machine Learning,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 6, pp. 207–215, 2022.
- [12] A. Saiful, “Prediksi Harga Rumah Menggunakan Web Scrapping dan Machine Learning Dengan Algoritma Linear Regression,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 41–50, 2021.
- [13] I. M. Muhamad, S. A. Wardana, A. Wanto, and A. P. Windarto, “Algoritma Machine Learning untuk penentuan Model Prediksi Produksi Telur Ayam Petelur di Sumatera,” *J. Informatics ...*, vol. 1, no. 4, pp. 126–134, 2022.
- [14] F. M. Palechor and A. de la H. Manotas, “Estimation of obesity levels based on eating habits and physical condition Data Set,” *UCI*, 2019.
- [15] Z. Peng and N. Niu, “Co-AI: A Colab-Based Tool for Abstraction Identification,” *Proc. IEEE Int. Conf. Requir. Eng.*, pp. 420–421, 2021.
- [16] E. Camizuli and E. J. Carranza, “Exploratory Data Analysis (EDA),” *Encycl. Archaeol. Sci.*, no. 3, pp. 1–7, 2018.
- [17] Z. Noshad *et al.*, “Fault detection in wireless sensor networks through the random forest classifier,” *Sensors (Switzerland)*, vol. 19, no. 7, pp. 1–21, 2019.
- [18] R. G. “A Study to Find Facts Behind Preprocessing on Deep Learning Algorithms,” *J. Innov. Image Process.*, vol. 3, no. 1, pp. 66–74, 2021.
- [19] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, “Perbandingan Klasifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 3, pp. 432–437, 2021.
- [20] I. Sutoyo, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik,” *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 217, 2018.
- [21] M. I. Gunawan, D. Sugiarto, and I. Mardianto, “Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search pada Algoritma Logistic Regression,” *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 280, 2020.
- [22] M. K. Anam, B. N. Pikir, and M. B. Firdaus, “Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemerintah,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 139–150, 2021.
- [23] A. Sulistiyo, “Penentuan Jurusan Sekolah Menengah Atas Menggunakan Metode K-Nearest Neighbor Classifier Pada SMAN 16 Semarang,” *Fasilkom Udinus*, vol. 1, no. 1, pp. 1–5, 2014.

