

## KLASIFIKASI BMI (*BODY MASS INDEX*) BERDASARKAN TINGGI DAN BERAT BADAN MENGGUNAKAN *LOGISTIC REGRESSION*

Dery Ferdika Oktoriansah<sup>1\*</sup>

(Matematika, FMIPA, Universitas Negeri Malang, [dery.ferdika.2003126@students.um.ac.id](mailto:dery.ferdika.2003126@students.um.ac.id)<sup>1</sup>)

\*Email : [dery.ferdika.2003126@students.um.ac.id](mailto:dery.ferdika.2003126@students.um.ac.id)

### Abstrak

*BMI (Body Mass Index) atau juga disebut Indeks Massa Tubuh adalah suatu indikator untuk memantau status gizi orang dewasa, khususnya berkaitan dengan kekurangan dan kelebihan berat badan (obesitas). Logistic Regression adalah regresi yang menguji apakah terdapat probabilitas terjadinya variabel dependen dapat diprediksi oleh variabel independen. Penelitian ini menggunakan data 500 orang dengan tinggi dan beratnya masing-masing serta indeks BMI mulai dari indeks ke 1 sampai indeks ke 5. Didapat bahwa keakuratan model yang dihasilkan sebesar 78%.*

**Kata kunci:** BMI, Regresi, Logistik, Model

### PENDAHULUAN

Menurut Goldberg & Holland (1988), Machine learning dapat didefinisikan sebagai aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan prediksi di masa yang akan datang [1]. Salah satu metode dari *Machine Learning* yang sering dipakai adalah *Supervised Learning*. *Supervised Learning* adalah suatu algoritma *Machine Learning* di mana data yang tersedia sudah memiliki label dan output-nya sudah jelas.

Salah satu contoh model dari *Supervised Learning* adalah Regresi Logistik. Meskipun Namanya adalah Regresi, namun Regresi Logistik sejatinya termasuk klasifikasi. Di artikel ini, kami meneliti tentang BMI (*Body Mass Index*) dari data 500 pasien. Tujuan dari penelitian ini adalah mengembangkan suatu model *Machine Learning* yang dapat memprediksi klasifikasi BMI dengan akurasi dan presisi yang tinggi. Harapannya, setelah kami mengembangkan model ini, para petugas Kesehatan dapat menggunakan model ini untuk memprediksi klasifikasi BMI dengan akurasi dan presisi yang tinggi sehingga pasien dapat terklasifikasi dengan baik dan benar.

### METODE

Penelitian ini menggunakan salah satu algoritma *Machine Learning* yaitu *Supervised Learning*, di mana algoritma tersebut memiliki salah satu model bernama Regresi Logistik. Regresi Logistik adalah suatu model *Machine Learning* untuk mendeskripsikan hubungan antara variabel terikat yang memiliki dua kategori atau lebih dengan satu atau lebih peubah bebas berskala kategori atau kontinu [2].

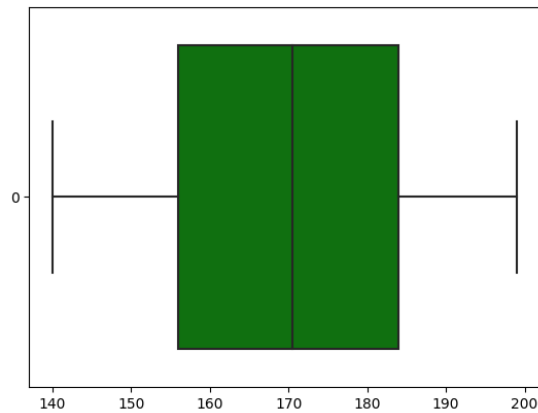
Model yang akan digunakan dalam penelitian ini adalah Regresi Logistik tanpa *hyperparameter tuning*. Target dari penelitian ini adalah 500 orang dengan berat, tinggi, gender, dan indeks BMI-nya masing-masing. Data penelitian ini diambil dari situs *Kaggle* serta platform atau aplikasi yang kami gunakan dalam penelitian ini adalah Google Colaboratory. Adapun teknik analisis yang digunakan ada 4 tahap, yaitu tahap *Data Collecting*, *Data Preparation*, *Modelling*, dan *Model Evaluation*..

Tahap *Data Collecting* adalah tahap di mana kami mengambil data dari *Kaggle* serta meng-import library yang dibutuhkan seperti *library Pandas*, *Numpy*, *Seaborn*, *Matplotlib*, dan lain-lain. Tahap *Data Preparation* adalah tahap di mana kami membersihkan data dari nilai kosong atau *Missing Values Handling*, mengatasi *outliers* atau data yang terpaut jauh dari data lainnya, dan melakukan *Exploratory Data Analysis* untuk mencari hubungan antara variabel-variabel yang ada.

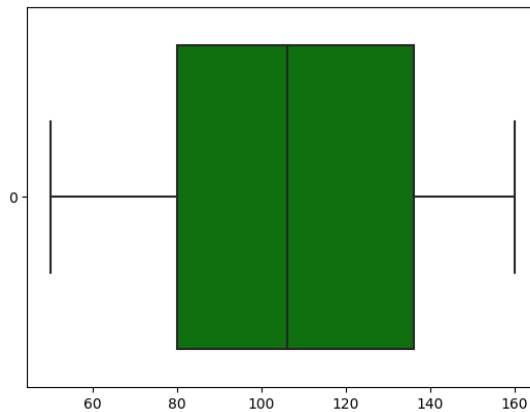
Tahap *Modelling* adalah tahap di mana kami mengimplementasikan model Regresi Logistik ke dalam data yang telah di-split sebanyak 80% untuk data *training* (data pelatihan) dan 20% untuk data *testing* (data uji). Tahap *Model Evaluation* adalah tahap untuk melakukan evaluasi model seperti mengecek ulang akurasi, presisi, *recall*, dan *confusion matrix* (matriks kebingungan)

## HASIL DAN PEMBAHASAN

Dari tahap *Data Collecting* dan *Data Preparation* didapat bahwa Tinggi dan Berat pasien tidak mengandung *outlier* serta tidak terdapat *missing values*. Terlihat jelas di visualisasi boxplot di bawah ini bahwa data tinggi badan dan berat badan pasien terdistribusi normal.



Gambar 1. Visualisasi Kolom *Height* Menggunakan Boxplot



Gambar 2. Visualisasi Kolom *Weight* Menggunakan Boxplot

Setelah melalui tahap *Data Collecting* dan *Data Preparation*, maka tahap selanjutnya adalah tahap *Modelling*, yaitu memodelkan data pasien dengan model Regresi Logistik. Sebelum melakukan *Modelling*, dataset pasien yang telah bersih tersebut harus dibagi menjadi 2 bagian yaitu data *train* dan data *test*. Tujuan membagi dataset menjadi 2 agar tidak terjadi bias pada hasil model. Kami membagi dataset menjadi 80% untuk data *train* dan 20% untuk data *test*. Setelah membagi dataset menjadi data *train* dan data *test*, maka selanjutnya adalah mengimplementasikan model Regresi Logistik ke dataset yang telah dibagi tersebut. Kami meng-*import* Pustaka atau *library* dari Regresi Logistik itu sendiri. Data yang dikenai model Regresi Logistik adalah data *train* karena kami akan melatih model dan nantinya akan diuji dengan data *test* untuk dicocokkan melalui akurasi, presisi, dan *recall*-nya.

```

from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

# Buat objek model Logistic Regression
model = LogisticRegression()

# Latih model dengan data latih
model.fit(X_train, y_train)

```

▼ LogisticRegression  
LogisticRegression()

Gambar 3. Tahap *Modelling Data Train* menggunakan Pustaka sklearn.

Tahap terakhir adalah tahap *Model Evaluation* yaitu mengevaluasi model yang telah dibuat. Tujuan dari tahap ini adalah untuk mengetahui seberapa besar akurasi model dalam menjelaskan permasalahan tersebut. Didapat bahwa akurasi dari model Regresi Logistik adalah 0.78 atau 78%. Artinya model mampu menjelaskan 78% dari permasalahan BMI tersebut, 22% sisanya dijelaskan oleh variable lain. Dari *confusion matrix* atau matriks kebingungan sendiri, didapat hasil bahwa ada 16 sample yang gagal dijelaskan oleh model.

```

Akurasi: 0.78
Presisi: 0.7900802355539197
Recall: 0.78
F1-score: 0.7823855092276145
Confusion Matrix:
[[ 1  0  0  0  0  0]
 [ 0  2  2  0  0  0]
 [ 0  2 16  0  0  0]
 [ 0  0  2  5  1  0]
 [ 0  0  0  6 20  4]
 [ 0  0  0  0  5 34]]

```

Gambar 4. Tahap *Model Evaluation*.

## PENUTUP

Kesimpulan dari penelitian ini adalah model Regresi Logistik memiliki akurasi sebesar 0.78 atau 78% yang artinya model mampu menjelaskan 78% dari permasalahan tersebut. Keterbatasan penelitian ini adalah kurangnya peningkatan parameter atau *hyperparameter tuning* demi meningkatkan performa model dan kurangnya variasi algoritma *Machine Learning* yang digunakan. Saran yang dapat kami sampaikan untuk penelitian selanjutnya adalah sebaiknya gunakan *hyperparameter tuning* untuk meningkatkan performa akurasi model serta menggunakan variasi model *Machine Learning* yang lain seperti Linear Regression, *Decision Tree*, *Random Forest*, dan lain-lain.

## DAFTAR RUJUKAN

- [1] Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. 3 (2): 95-99.
- [2] Hosmer, D.W., dan S. Lemeshow. 2000. *Applied Logistic Regression*. Edisi ke-2. John Wiley and Sons Inc, Canada.