

**PERBANDINGAN KINERJA ALGORITMA RIDGE DAN LASSO  
REGRESSION DALAM PREDIKSI PERSENTASE  
LEMAK TUBUH**

**SKRIPSI**

**AMMAR WAHYUDI**

**201401095**



**PROGRAM STUDI ILMU KOMPUTER  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA**

**MEDAN**

**2025**

**PERBANDINGAN KINERJA ALGORITMA RIDGE DAN LASSO  
REGRESSION DALAM PREDIKSI PERSENTASE  
LEMAK TUBUH**

**SKRIPSI**

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah

Sarjana Ilmu Komputer

**AMMAR WAHYUDI**

**201401095**



**PROGRAM STUDI ILMU KOMPUTER**

**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

**UNIVERSITAS SUMATERA UTARA**

**MEDAN**

**2025**

**PERSETUJUAN**

Judul : PERBANDINGAN KINERJA ALGORITMA RIDGE  
DAN LASSO REGRESSION DALAM PREDIKSI  
PERSENTASE LEMAK TUBUH

Kategori : SKRIPSI

Nama : AMMAR WAHYUDI

Nomor Induk Mahasiswa : 201401095

Program : SARJANA (S1) ILMU KOMPUTER

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA


Medan, 10 Januari 2025

Komisi Pembimbing :

Pembimbing II

Pembimbing I

  
Prof. Dr. Syahril Efendi S.Si., M.I.T.  
NIP. 196711101996021001

  
Desilia Selvida S.Kom., M.Kom.  
NIP. 198912052024062002

Diketahui/Disetujui Oleh

Program Studi S1 Ilmu Komputer



Dr. Amalia, S.T., M.T.

NIP. 197812212014042001

**PERNYATAAN****PERBANDINGAN KINERJA ALGORITMA RIDGE DAN LASSO  
REGRESSION DALAM PREDIKSI PERSENTASE  
LEMAK TUBUH****SKRIPSI**

Saya mengakui bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 10 Januari 2025



Ammar Wahyudi

201401095

## UCAPAN TERIMA KASIH

Peneliti mengucapkan rasa syukur kepada Allah SWT dengan mengucapkan, Alhamdulillah. Dengan kasih sayang dan karunia-Nya, peneliti dapat menyusun skripsi ini. Penulis juga menyampaikan shalawat dan salam kepada junjungan Nabi Muhammad SAW. "Allahumma sholli 'ala ta'ala Muhammad wa'ala ta'ala Muhammad".

Peneliti mengutarakan penghargaan bagi seluruh pihak karena senantiasa menyalurkan semangat serta panduan terhadap peneliti. Kalimat terima kasih peneliti sampaikan untuk:

1. Allah SWT. Atas semua rahmat dan petunjuk-Nya, penulis mampu berada di posisi sekarang ini.
2. Bapak Prof. Dr. Muryanto Amin, S.Sos, M.Si. selaku Rektor Universitas Sumatera Utara.
3. Ibu Dr. Maya Silvi Lydia, M.Sc. selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
4. Ibu Dr. Amalia S.T., M.T. selaku Ketua Prodi Ilmu Komputer Universitas Sumatera Utara
5. Ibu Desilia Selvida S.Kom., M.Kom. selaku Dosen Bimbingan I penulis.
6. Bapak Prof. Dr. Syahril Efendi S.Si., M.I.T. selaku Dosen Bimbingan II dan Dosen Penasehat Akademik penulis.
7. Seluruh staf pengajar dan pegawai Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
8. Orang tua peneliti yaitu atas nama Mariani dan H. Mahyudin, anggota keluarga seperti kakak peneliti yaitu atas nama Maryuani, Irma Rahayu, Junita Wahyu, dan Siti Hartina, abang ipar peneliti yaitu Agung Rizky dan Sunandi Effendi, keponakan peneliti yaitu Aris, Hakim, Fahrul, Umar, Ayra, Yumna, dan anggota keluarga lainnya karena telah menyampaikan doa, panduan, serta antusias kepada peneliti.
9. Orang terdekat, sahabat, serta teman peneliti yaitu atas nama Sutan, Parhan, Topan, Jessica, Yunisa, Fauzi, Yogi, Jimmi, Irfan, Sonia, Farel, Feli, Disa, Nanda, dan teman-teman lainnya yang belum dapat disebutkan oleh peneliti.



10. Dan semua orang yang secara langsung maupun tidak langsung berkontribusi namun peneliti belum dapat menyebutkan namanya satu persatu.
11. Terakhir, penulis ingin mengucapkan banyak rasa terima kasih kepada diri penulis sendiri yang telah mampu menyelesaikan skripsi ini dan berdoa agar dimasa depan penulis diberi kesehatan dan menjadi pribadi yang lebih baik dan dapat berguna kepada orang sekitar.

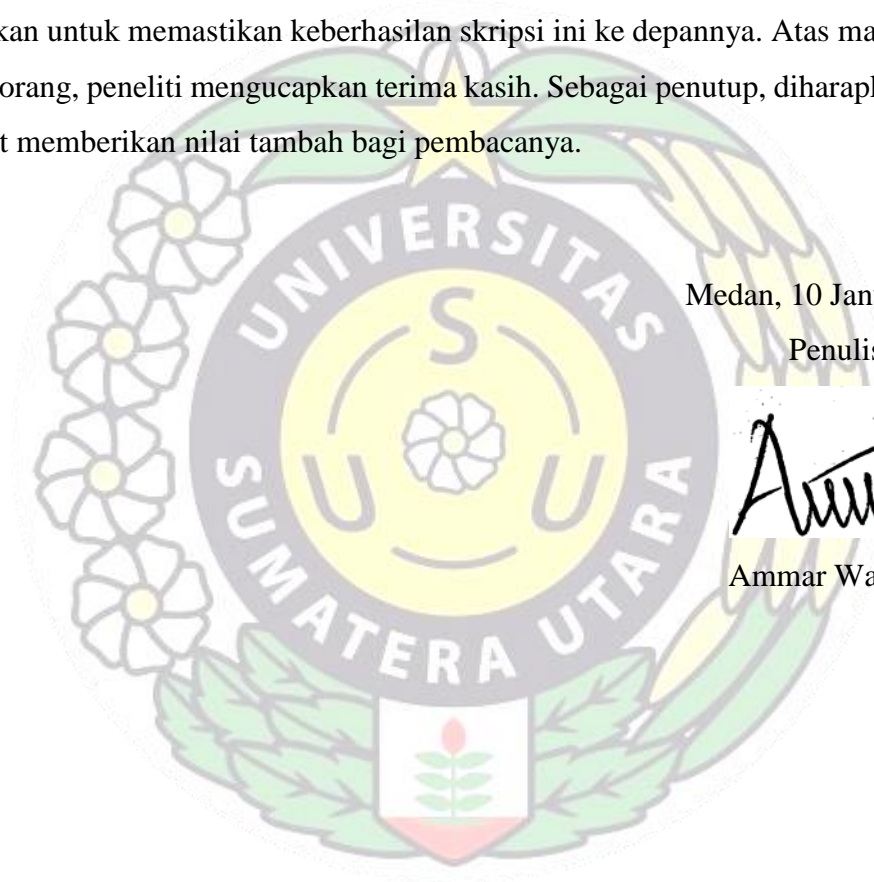
Dengan segala keterbatasan, pendapat, dan masukan dari seluruh orang yang sangat diharapkan untuk memastikan keberhasilan skripsi ini ke depannya. Atas masukan dari seluruh orang, peneliti mengucapkan terima kasih. Sebagai penutup, diharapkan skripsi ini dapat memberikan nilai tambah bagi pembacanya.

Medan, 10 Januari 2025

Penulis,



Ammar Wahyudi



## ABSTRAK

### PERBANDINGAN KINERJA ALGORITMA RIDGE DAN LASSO REGRESSION DALAM PREDIKSI PERSENTASE LEMAK TUBUH

Persentase lemak tubuh atau *Body Fat Percentage* (BFP) adalah pengukuran yang tepat untuk diagnosis obesitas, faktor utama yang menyebabkan obesitas adalah ketidakseimbangan energi antara pengeluaran dan asupan kalori. Orang yang mengalami obesitas lebih mungkin mengalami diabetes, penyakit jantung, stroke, dan masalah muskuloskeletal. Menurut *World Health Organization* (WHO), lebih dari 1,9 miliar orang berusia di atas 18 tahun mengalami kelebihan berat badan pada tahun 2016, dan lebih dari 650 juta di antaranya mengalami obesitas. Nilai BFP sangat penting untuk menentukan kemungkinan seseorang mengalami obesitas, tetapi mengukurnya dapat menjadi tantangan, mahal, dan menyakitkan. Untuk memprediksi nilai BFP seseorang berdasarkan ukuran tubuh mereka, penelitian ini menggunakan teknik penambangan data, khususnya algoritma Ridge Regression dan LASSO Regression. Kedua algoritma ini menggunakan teknik regularisasi yaitu dengan menambahkan penalti pada nilai koefisien ke arah nol. Data yang digunakan berjumlah 250 baris, yang dibagi menjadi dua bagian berdasarkan persentase 80:20 dari total data. Sebanyak 200 baris data (80%) digunakan sebagai data latih, sedangkan 50 baris data (20%) digunakan sebagai data uji. Hasil penelitian ini menunjukkan bahwa algoritma LASSO Regression lebih unggul untuk keseluruhan matrik evaluasi, dengan nilai MSE sebesar 22.56, MAE sebesar 3.64, R-squared sebesar 0.60, dan Adjusted R-squared sebesar 0.44, berbanding terbalik dengan Ridge Regression dengan nilai MSE hanya sebesar 22.89, MAE sebesar 3.80, R-squared sebesar 0.59, dan Adjusted R-squared sebesar 0.43.

**Kata Kunci:** LASSO Regression, Lemak Tubuh, Prediksi, Ridge Regression.

## ABSTRACT

### PERFORMANCE COMPARISON OF RIDGE AND LASSO REGRESSION ALGORITHMS IN BODY FAT PERCENTAGE PREDICTION

Body Fat Percentage (BFP) is an accurate measurement for diagnosing obesity, the main factor causing obesity is an energy imbalance between calorie expenditure and intake. People who are obese are more likely to experience diabetes, heart disease, stroke, and musculoskeletal problems. According to the World Health Organization (WHO), more than 1.9 billion people aged 18 and over were overweight in 2016, and more than 650 million of them were obese. The BFP value is very important for determining the likelihood of someone being obese, but measuring it can be challenging, expensive, and painful. To predict a person's BFP value based on their body size, this study used data mining techniques, specifically the Ridge Regression and LASSO Regression algorithms. Both of these algorithms use regularization techniques, namely by adding a penalty to the coefficient value towards zero. The data used amounted to 250 rows, which were divided into two parts based on the percentage of 80:20 of the total data. A total of 200 rows of data (80%) were used as training data, while 50 rows of data (20%) were used as test data. The results of this study indicate that the LASSO Regression algorithm is superior for the entire evaluation matrix, with an MSE value of 22.56, MAE of 3.64, R-squared of 0.60, and Adjusted R-squared of 0.44, inversely proportional to Ridge Regression with an MSE value of only 22.89, MAE of 3.80, R-squared of 0.59, and Adjusted R-squared of 0.43.

**Keywords:** Body Fat, LASSO Regression, Prediction, Ridge Regression.



## DAFTAR ISI

|                                               |      |
|-----------------------------------------------|------|
| PERSETUJUAN .....                             | iii  |
| PERNYATAAN .....                              | iv   |
| UCAPAN TERIMA KASIH.....                      | v    |
| ABSTRAK.....                                  | vii  |
| ABSTRACT.....                                 | viii |
| DAFTAR ISI.....                               | ix   |
| DAFTAR GAMBAR .....                           | xi   |
| DAFTAR TABEL.....                             | xii  |
| BAB I PENDAHULUAN.....                        | 1    |
| 1.1. Latar Belakang .....                     | 1    |
| 1.2. Rumusan Masalah .....                    | 3    |
| 1.3. Batasan Masalah.....                     | 3    |
| 1.4. Tujuan Penelitian.....                   | 3    |
| 1.5. Manfaat Penelitian.....                  | 4    |
| 1.6. Metodologi Penelitian .....              | 4    |
| 1.7. Penelitian Relevan.....                  | 5    |
| 1.8. Sistematika Penulisan.....               | 6    |
| BAB II LANDASAN TEORI.....                    | 8    |
| 2.1. Obesitas dan Lemak Tubuh.....            | 8    |
| 2.2. Data Mining.....                         | 9    |
| 2.3. Machine Learning .....                   | 9    |
| 2.4. Multiple Linear Regression.....          | 10   |
| 2.5. Ridge Regression.....                    | 11   |
| 2.6. LASSO Regression .....                   | 12   |
| BAB III ANALISIS DAN PERANCANGAN SISTEM ..... | 13   |

|                                                              |           |
|--------------------------------------------------------------|-----------|
| 3.1. Analisis Sistem.....                                    | 13        |
| 3.1.1. Analisis Masalah.....                                 | 13        |
| 3.1.2. Analisis Requirements .....                           | 13        |
| 3.1.3. Arsitektur Sistem .....                               | 15        |
| 3.2. Pemodelan Sistem .....                                  | 16        |
| 3.2.1. Use Case Diagram .....                                | 16        |
| 3.2.2. Activity Diagram .....                                | 17        |
| 3.3. Flowchart.....                                          | 19        |
| 3.3.1. Flowchart Ridge Regression.....                       | 19        |
| 3.3.2. Flowchart LASSO Regression.....                       | 20        |
| 3.3.3. Flowchart Sistem .....                                | 21        |
| 3.4. Perancangan Interface .....                             | 21        |
| 3.4.1. Interface Halaman Tampilan Sistem .....               | 22        |
| <b>BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM .....</b>        | <b>24</b> |
| 4.1. Implementasi Sistem .....                               | 24        |
| 4.1.1. Halaman Tampilan Sistem.....                          | 24        |
| 4.1.2. Halaman Tampilan Hasil Prediksi .....                 | 25        |
| 4.2. Pra-Pemrosesan Dataset .....                            | 25        |
| 4.3. Implementasi Algoritma Ridge dan LASSO Regression ..... | 29        |
| 4.4. Pengujian Model .....                                   | 34        |
| 4.5. Hasil Pengujian .....                                   | 40        |
| <b>BAB V KESIMPULAN DAN SARAN .....</b>                      | <b>41</b> |
| 5.1. Kesimpulan.....                                         | 41        |
| 5.2. Saran .....                                             | 41        |
| <b>DAFTAR PUSTAKA .....</b>                                  | <b>42</b> |

## DAFTAR GAMBAR

|                                                        |    |
|--------------------------------------------------------|----|
| <b>Gambar 3.1</b> Arsitektur Sistem .....              | 15 |
| <b>Gambar 3.2</b> Use Case Diagram .....               | 17 |
| <b>Gambar 3.3</b> Activity Diagram .....               | 18 |
| <b>Gambar 3.4</b> Flowchart Ridge Regression.....      | 19 |
| <b>Gambar 3.5</b> Flowchart LASSO Regression .....     | 20 |
| <b>Gambar 3.6</b> Flowchart Sistem .....               | 21 |
| <b>Gambar 3.7</b> Interface Halaman Home.....          | 22 |
| <b>Gambar 3.8</b> Interface Halaman Predict .....      | 22 |
| <b>Gambar 4.1</b> Halaman Home .....                   | 24 |
| <b>Gambar 4.2</b> Halaman Predict .....                | 25 |
| <b>Gambar 4.3</b> Tampilan Hasil Prediksi .....        | 25 |
| <b>Gambar 4.4</b> Import Dataset .....                 | 26 |
| <b>Gambar 4.5</b> Cek Missing Values.....              | 26 |
| <b>Gambar 4.6</b> Cek Duplicated Data.....             | 26 |
| <b>Gambar 4.7</b> Anomali Data Bodyfat bernilai 0..... | 27 |
| <b>Gambar 4.8</b> Outlier.....                         | 27 |
| <b>Gambar 4.9</b> Konversi Satuan Weight.....          | 27 |
| <b>Gambar 4.10</b> Konversi Satuan Height.....         | 27 |
| <b>Gambar 4.11</b> Perhitungan BMI .....               | 27 |
| <b>Gambar 4.12</b> Dataset Model.....                  | 28 |
| <b>Gambar 4.13</b> Train Test Split .....              | 28 |
| <b>Gambar 4.14</b> Standarisasi Data .....             | 28 |
| <b>Gambar 4.15</b> Grid Search .....                   | 29 |
| <b>Gambar 4.16</b> Train Model Ridge dan LASSO .....   | 29 |
| <b>Gambar 4.17</b> Sampel Data.....                    | 31 |
| <b>Gambar 4.18</b> Standarisasi Sampel Data .....      | 31 |

**DAFTAR TABEL**

|                                                           |    |
|-----------------------------------------------------------|----|
| <b>Tabel 2.1</b> Klasifikasi BMI.....                     | 8  |
| <b>Tabel 4.1</b> Nilai Koefisien .....                    | 30 |
| <b>Tabel 4.2</b> Hasil Prediksi .....                     | 34 |
| <b>Tabel 4.3</b> Matriks Evaluasi.....                    | 36 |
| <b>Tabel 4.4</b> Perbandingan Nilai Matrik Evaluasi ..... | 40 |



# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Salah satu faktor penting yang mempengaruhi kualitas hidup seseorang adalah kesehatan fisiknya. BFP (*Body Fat Percentage*), sebutan lain untuk persentase lemak tubuh, adalah metrik yang digunakan untuk menilai kesehatan seseorang. Mempertahankan persentase lemak tubuh yang sehat sangat penting untuk pencegahan penyakit dan fungsi tubuh yang optimal. Teknik pengukuran lemak tubuh idealnya harus akurat, praktis, dan tersedia secara luas.

Menurut World Health Organization (WHO), kelebihan berat badan atau obesitas memiliki tingkat kematian global yang lebih tinggi daripada kekurangan berat badan. Lebih dari 1,9 miliar orang berusia di atas 18 tahun mengalami kelebihan berat badan pada tahun 2016, dan lebih dari 650 juta di antaranya mengalami obesitas. 13% individu di seluruh dunia mengalami obesitas pada tahun yang sama (11% pria dan 15% wanita), sementara 39% orang dewasa berusia di atas 18 tahun, 40% wanita, dan 39% pria mengalami kelebihan berat badan. Prevalensi obesitas di seluruh dunia telah meningkat tiga kali lipat sejak 1975. Obesitas dan kelebihan berat badan terutama disebabkan oleh ketidakseimbangan energi antara asupan dan pengeluaran kalori. Risiko kesehatan umum yang terkait dengan kondisi ini meliputi diabetes, gangguan muskuloskeletal, dan gangguan kardiovaskular, khususnya penyakit jantung dan stroke (World Health Organization, 2024).

Persentase lemak tubuh (BFP) adalah pengukuran yang tepat untuk diagnosis obesitas. Ada berbagai cara untuk menghitung persentase lemak tubuh. Pada metode *Anthropometry*, dilakukan pengukuran seperti tinggi badan, berat badan, dan lingkar pinggang. *Dual-energy X-ray Absorptiometry* (DXA) menggunakan sinar-X untuk mengukur kepadatan tulang dan lemak tubuh, metode ini memiliki akurasi yang baik. Namun, DXA tidak cocok untuk dilakukan secara rutin karena dosis radiasi. *Skinfold*



*thickness* atau pengukuran ketebalan lipatan kulit meskipun mudah dan murah, metode ini dipengaruhi oleh beberapa faktor seperti cengkeraman praktisi, tekanan yang diberikan kaliper serta jenis dan merek kaliper yang digunakan. Dalam penelitian ini baik pengukuran *Anthropometry* dan *Skinfold thickness* dilakukan oleh staf dengan sertifikasi level 1 dari *International Society for the Advancement of Kinanthropometry* (ISAK) (Huang et al., 2023).

Dalam kondisi ideal, metode pengukuran lemak tubuh seharusnya mudah diakses, tidak invasif, dan memiliki akurasi yang tinggi. Namun, metode tradisional yang ada saat ini seringkali tidak memenuhi semua kriteria tersebut. Hal ini mengakibatkan keterlambatan dalam diagnosis dini lemak tubuh yang dapat meningkatkan resiko penyakit kronis seperti penyakit jantung, stroke, dan diabetes.

Sebagai solusi dalam prediksi persentase lemak tubuh, penelitian ini mengusulkan penggunaan algoritma *machine learning*, yaitu *Ridge Regression* dan *LASSO Regression* berdasarkan dataset dari Kaggle. Regresi *LASSO* (*Least Absolute Shrinkage and Selection Operator*) unggul dalam seleksi fitur, dengan kemampuan mengidentifikasi fitur-fitur yang paling signifikan dan mengeliminasi variabel yang tidak relevan, sehingga membuat model lebih sederhana dan mudah dipahami. Di sisi lain, *Ridge Regression* menawarkan keunggulan dalam mengatasi *overfitting* dan masalah multikolinearitas, terutama ketika variabel independen berkorelasi tinggi (Ren, 2024). Namun, hasil penelitian tentang efektivitas kedua algoritma ini tidak selalu konsisten. Penelitian (Ren, 2024) menemukan bahwa *LASSO Regression* memiliki kinerja yang lebih baik, sedangkan penelitian (Thevaraja et al., 2019) menunjukkan bahwa *Ridge Regression* lebih unggul. Ren juga menyatakan bahwa performa algoritma dapat berbeda tergantung oleh jenis dataset yang digunakan, sehingga perbandingan antara *Ridge* dan *LASSO Regression* dalam penelitian ini diharapkan akan memperdalam pemahaman tentang algoritma terbaik untuk memperkirakan persentase lemak tubuh.

Penelitian ini menghasilkan model prediksi persentase lemak tubuh berbasis *website* menggunakan algoritma terbaik yang didapat dari hasil perbandingan kinerja antara algoritma *Ridge* dan *LASSO Regression*. Diharapkan hasil penelitian ini dapat membantu profesional kesehatan dalam melakukan diagnosis awal yang lebih cepat dan efisien, dan menjadi referensi untuk penelitian lebih lanjut tentang penggunaan *machine*

*learning* untuk memprediksi kondisi kesehatan serta perkembangan teknologi di bidang kesehatan.

## 1.2. Rumusan Masalah

Persentase lemak tubuh (BFP) merupakan indikator penting yang berkaitan dengan risiko penyakit kronis seperti obesitas, diabetes, dan penyakit kardiovaskular. Meski pengukurannya dapat dilakukan secara manual, metode ini kurang efisien dari segi waktu dan biaya. Dalam upaya meningkatkan akurasi dan efisiensi prediksi persentase lemak tubuh, penggunaan algoritma *machine learning* menjadi solusi yang relevan. *Ridge* dan *LASSO Regression* dipilih karena keduanya mampu menangani data dengan jumlah variabel yang banyak dan masalah multikolinearitas. Oleh karena itu, penting untuk mengetahui bagaimana kinerja kedua algoritma tersebut dalam memprediksi persentase lemak tubuh, dan algoritma mana yang memberikan hasil prediksi lebih akurat.

## 1.3. Batasan Masalah

Beberapa batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Menggunakan algoritma *Ridge* dan *LASSO Regression*.
2. Variabel independen yang digunakan ada 14, yaitu usia, berat badan, tinggi badan, BMI, dan lingkaran dari leher, dada, perut, pinggul, paha, lutut, pergelangan kaki, bicep, lengan bawah, serta pergelangan tangan.
3. Matrik evaluasi yang digunakan adalah MAE (*Mean Absolute Error*), MSE (*Mean Squared Error*),  $R^2$  (*R-squared*), dan *Adjusted R-squared*.
4. Data didapatkan dari *website Kaggle* dan fokus pada data Pria.
5. Bahasa Pemrograman yang digunakan adalah Python, HTML, CSS, dan *Javascript*.
6. Program dirancang berbasis *Website*.

## 1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan sistem yang dapat memprediksi persentase lemak tubuh dengan memanfaatkan algoritma terbaik antara *Ridge* dan *LASSO Regression* dengan menginputkan variabel.

### 1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Dapat melakukan prediksi persentase lemak tubuh menggunakan algoritma terbaik diantara *Ridge* dan *LASSO regression* dengan cepat dan akurat.
2. Dapat dijadikan sebagai referensi dalam pengembangan penelitian selanjutnya mengenai sistem prediksi persentase lemak tubuh menggunakan algoritma *Ridge* dan *LASSO regression*.
3. Dapat membantu profesional kesehatan dalam melakukan diagnosis awal yang lebih cepat dan efisien.

### 1.6. Metodologi Penelitian

Metodologi penelitian untuk riset ini adalah sebagai berikut:

1. Studi Pustaka  
Pada tahap ini, penulis mulai dengan mengumpulkan referensi dan meninjau literatur berupa jurnal dan penelitian yang berkaitan dengan lemak tubuh serta penerapan algoritma *Ridge regression* dan *LASSO regression*.
2. Pengumpulan Data  
Pada tahap ini, tugas penulis adalah mempelajari data yang dibutuhkan untuk memperkirakan lemak tubuh. Informasi ini dikumpulkan dari *Body Fat Prediction Dataset* di situs web *Kaggle*, yang mencakup estimasi lemak tubuh dan berbagai pengukuran lingkaran tubuh untuk 252 pria.
3. Analisa dan Perancangan Sistem  
Pada tahap ini, penulis menganalisis persyaratan fungsional dan nonfungsional untuk sistem dan membuat diagram alir, desain antarmuka interaktif, diagram kasus penggunaan, dan diagram aktivitas.
4. Implementasi Sistem  
Sesuai dengan diagram alir yang telah dibuat, penulis menggunakan bahasa pemrograman Python, HTML, CSS, dan Javascript untuk membuat sistem.
5. Pengujian Sistem  
Pada tahap ini, algoritma terbaik antara *Ridge* dan *LASSO Regression* digunakan untuk menilai kemampuan sistem yang diusulkan dalam memperkirakan persentase lemak tubuh.



## 6. Dokumentasi Sistem

Pada tahap ini, penulis melakukan dokumentasi dan penulisan laporan dalam bentuk penulisan skripsi.

### 1.7. Penelitian Relevan

Berikut ini beberapa penelitian yang berkaitan dengan penelitian ini.

1. Berdasarkan penelitian (Ren, 2024) dengan judul “*Comparison and analysis of the accuracy of Lasso regression, Ridge regression and Elastic Net regression models in predicting students' teaching quality achievement*”, berdasarkan penelitian yang telah dilakukan menyimpulkan bahwa *LASSO* merupakan model dengan performa terbaik pada dataset yang digunakan, diikuti oleh *Elastic Net* dan *Ridge* dengan performa terburuk, hal ini disebabkan karena terdapat korelasi positif pada dataset, peneliti juga menyimpulkan bahwa hasil dapat berbeda pada dataset yang lain.
2. Berdasarkan penelitian (Hasanah & Nurhopipah, 2023) dengan judul “Prediksi Persentase Body Fat Menggunakan Algoritma CART dan M5”, dalam penelitian ini, Persentase Lemak Tubuh (BFP) digunakan sebagai metrik yang tepat untuk mendiagnosis obesitas. Data diekstraksi dari *Kaggle* dengan menghilangkan variabel *density* karena korelasinya yang negatif. *Mean Absolute Error* (MAE) digunakan untuk menilai kinerja model, dan algoritma M berkinerja lebih baik dengan nilai korelasi 0,86 dan MAE 3,86.
3. Berdasarkan penelitian (Enwere et al., 2023) dengan judul “*Comparative Analysis of Ridge, Bridge and Lasso Regression Models In the Presence of Multicollinearity*”, berdasarkan penelitian yang telah dilakukan menyimpulkan bahwa ketiga algoritma yaitu, *Bridge*, *LASSO*, dan *Ridge Regression* dapat mengatasi multikolinearitas dan *overfitting*, meskipun tetap harus mempertimbangkan jenis datanya, parameter perbandingan yang digunakan yaitu, MSE, RMSE, VIF, AIC and BIC.
4. Berdasarkan penelitian (Sitanggang, 2022) dengan judul “Model Prediksi Obesitas Dengan Menggunakan Support Vector Machine”, Tujuan dari penelitian ini adalah untuk mempermudah penentuan tingkat kepadatan tubuh seseorang karena prosedur pengumpulan datanya sulit. Kinerja algoritma

dievaluasi menggunakan metode MAE dan RMSE, dan akurasi akhir adalah 71,80% dengan MSE 17,76.

5. Berdasarkan penelitian (Thevaraja et al., 2019) dengan judul “*Recent Developments in Data Science: Comparing Linear, Ridge and Lasso Regressions Techniques Using Wine Data*”, pada penelitian ini digunakan *cross validation* untuk menemukan nilai  $\lambda$  yang paling sesuai dengan dataset yang digunakan, didapatkan *Ridge* memiliki nilai error paling minimum diikuti *LASSO* dan terakhir *OLS*.

### 1.8. Sistematika Penulisan

Lima bab utama yang menyusun sistematika penulisan skripsi yang digunakan dalam penelitian ini adalah sebagai berikut:

#### **BAB I      PENDAHULUAN**

Latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, keuntungan penelitian, metodologi penelitian, penelitian yang relevan, dan sistematika penulisan semuanya dijelaskan dalam pendahuluan.

#### **BAB 2      LANDASAN TEORI**

Menyampaikan penjelasan mengenai lemak tubuh, *machine learning*, *data mining*, algoritma *ridge regression*, dan algoritma *LASSO regression*.

#### **BAB 3      ANALISIS DAN PERANCANGAN SISTEM**

Mengenai analisis sistem, pemodelan sistem, perancangan sistem, dan desain antarmuka dari penerapan algoritma untuk memprediksi persentase lemak tubuh.

#### **BAB 4      IMPLEMENTASI DAN PENGUJIAN SISTEM**

Proses penerapan sistem yang dirancang ke dalam praktik dan pengujiannya.



## **BAB 5 KESIMPULAN DAN SARAN**

Tentang temuan dan rekomendasi yang diperoleh dari temuan penelitian, yang dapat membantu memajukan penelitian masa depan.



## BAB II

### LANDASAN TEORI

#### 2.1. Obesitas dan Lemak Tubuh

Timbunan lemak berlebihan yang dapat membahayakan kesehatan merupakan ciri khas obesitas. Obesitas dapat mempengaruhi kesehatan tulang dan reproduksi, meningkatkan risiko beberapa jenis kanker, dan meningkatkan risiko penyakit jantung dan diabetes tipe 2. Selain itu, obesitas berdampak pada aktivitas sehari-hari, pola tidur, dan mobilitas. Tinggi dan berat badan seseorang diukur, dan indeks massa tubuh (IMT) mereka, yang merupakan berat badan (kg) / tinggi badan (m<sup>2</sup>), dihitung untuk mendiagnosis kelebihan berat badan dan obesitas.

**Tabel 2.1** Klasifikasi BMI

| Klasifikasi | BMI         |
|-------------|-------------|
| Underweight | <18.5       |
| Normal      | 18.5 – 22.9 |
| Overweight  | 23 – 24.9   |
| Obesity I   | 25 – 29.9   |
| Obesity II  | ≥30         |

Bahaya kelebihan berat badan atau obesitas semakin dikenal luas dan terdokumentasi dengan baik. Diperkirakan 5 juta kematian akibat penyakit tidak menular (PTM) termasuk kanker, diabetes, penyakit jantung, gangguan neurologis (gangguan saraf), kondisi pernapasan kronis, dan masalah pencernaan terjadi pada tahun 2019 akibat memiliki BMI yang lebih tinggi dari ideal (World Health Organization, 2024).

Selain itu, obesitas secara langsung terkait dengan lemak tubuh berlebih, yang meningkatkan risiko enam jenis kanker, termasuk kanker payudara, kanker usus, kanker esofagus, kanker pankreas, kanker ginjal, dan kanker endometrium. Pengukuran yang dapat diandalkan untuk mendiagnosis obesitas adalah persentase lemak tubuh (BFP),

yang merupakan pengukuran lemak tubuh total (Hasanah & Nurhopipah, 2023). Obesitas adalah masalah kesehatan masyarakat yang utama dengan prevalensi yang berkembang pesat dan risiko kesehatan yang serius. Peningkatan risiko sejumlah penyakit telah dikaitkan dengan obesitas, termasuk diabetes mellitus tipe 2, hipertensi, penyakit jantung koroner, stroke, infertilitas, osteoarthritis, penyakit hati dan kandung empedu, serta tumor tertentu (Ferenci & Kovács, 2018).

## 2.2. Data Mining

Data *mining* adalah proses menemukan pengetahuan dari gudang data. Pengetahuan ini dapat diklasifikasikan dalam berbagai aturan dan pola yang dapat membantu pengguna/organisasi untuk menganalisis data kolektif serta membantu proses mengambil keputusan berdasarkan data yang diprediksi. Jawad menyimpulkan bahwa data *mining* adalah sebuah konsep yang membantu menemukan informasi yang dibutuhkan dari gudang data yang besar dengan menggunakan berbagai teknik yang berbeda. Hal ini juga digunakan untuk menganalisis data masa lalu dan meningkatkan strategi masa depan (Jawad & Mughal, 2018).

Menemukan pola yang tidak teridentifikasi dalam data yang dikumpulkan merupakan tujuan dari sekelompok metode yang dikenal sebagai penambangan data. Penambangan data atau data mining digunakan untuk menemukan informasi atau kesimpulan yang dihipotesiskan. Berdasarkan penyaringan data untuk mencari tren atau ketidakteraturan. Ada lima fungsi dalam penambangan data: *Classification*, *Clustering*, *Association*, *Sequencing*, dan *Forecasting*. Dan memiliki tiga tujuan: *Explanatory*, *Confirmatory*, dan *Exploratory* (Zai, 2022).

## 2.3. Machine Learning

Pembelajaran (*learning*) sebagai proses umum adalah tentang memperoleh perilaku, nilai, pengetahuan, keterampilan, atau preferensi baru, atau memodifikasi perilaku, nilai, pengetahuan, keterampilan, atau preferensi yang sudah ada (Alzubi et al., 2018). Studi ilmiah tentang algoritma dan model statistik yang digunakan sistem komputer untuk menjalankan tugas tertentu tanpa pemrograman eksplisit dikenal sebagai pembelajaran mesin atau *machine learning* (Mahesh, 2020). Pembelajaran mesin (ML) dapat diterapkan pada berbagai bidang komputasi untuk merancang dan memprogram algoritma eksplisit dengan output berkinerja tinggi, misalnya, penyaringan *spam email*,

deteksi penipuan di jejaring sosial, perdagangan saham *online*, deteksi wajah & bentuk, diagnosis medis, prediksi lalu lintas, pengenalan karakter, rekomendasi produk, dan lain-lain.

Pembelajaran mesin (ML), yang didasarkan pada pembelajaran dan pengalaman berkelanjutan untuk memahami kompleksitas masalah dan kebutuhan untuk beradaptasi, diperlukan untuk memungkinkan komputer melakukan pekerjaan rumit tanpa campur tangan manusia. Dengan demikian, teknik pembelajaran mesin yang sesuai bervariasi berdasarkan sifat masalah. Dapat digunakan, termasuk *classification*, *anomaly detection*, *regression*, *clustering*, dan *reinforcement*. Dalam *supervised learning*, suatu algoritma belajar untuk merespons secara lebih akurat dengan membandingkan keluaran dengan masukan menggunakan kumpulan pelatihan contoh atau modul pelatihan yang berisi keluaran yang benar. *Unsupervised learning* adalah tentang mengenali pola yang tidak teridentifikasi dari data untuk mendapatkan aturan dari data tersebut. Teknik ini cocok digunakan dalam situasi ketika kategori data tidak diketahui. Di sini, data pelatihan tidak diberi label (Alzubi et al., 2018).

#### 2.4. Multiple Linear Regression

Regresi adalah metode statistik yang digunakan untuk menentukan hubungan yang ada antara variabel dependen dan variabel independen. Regresi linier berganda (*Multiple Linear Regression*) adalah kondisi dimana terdapat lebih dari satu variabel independen, dan seringkali menyebabkan multikolinieritas. Multikolinearitas adalah masalah yang dapat menyebabkan estimasi menjadi tidak stabil dan sulit menentukan efek dari setiap variabel independen terhadap variabel dependen. Untuk mengurangi multikolinearitas, dapat digunakan teknik regresi seperti *Ridge* dan *LASSO*, yang menggunakan regularisasi. Regularisasi adalah teknik menambahkan penalti pada model tertentu yang mengecilkan estimasi koefisien model ke arah nol untuk meminimalkan kesalahan. Teknik ini mengurangi estimasi varians dengan mengenalkan sedikit bias, sehingga meningkatkan akurasi prediksi (Enwere et al., 2023).

Rumus regresi linier berganda adalah:

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \dots + \delta_n w_n + \varepsilon \quad (2.1)$$



Dimana:

$v$  = variabel dependen

$w_1 w_2 w_3$  = variabel independen

$\delta_0$  = *y-intercept*

$\delta_1 \delta_2 \delta_3$  = nilai koefisien

$\varepsilon$  = *error*

Untuk mendapatkan estimasi  $\delta$  pada regresi linier berganda digunakan *sum of square error* (SSE). Berikut adalah rumus SSE:

$$\varepsilon^2 = \sum_{i=1}^n \left( v_i - \hat{\delta}_0 - \sum_{j=1}^p \hat{\delta}_j w_{ij} \right)^2 \quad (2.2)$$

SSE mengukur perbedaan antara data dan model estimasi, ketika SSE kecil berarti variansnya rendah dan ketika besar berarti variansnya tinggi.

## 2.5. Ridge Regression

Regresi *Ridge* diperkenalkan pertama kali oleh Hoerl dan R.W. Kennard pada tahun 1962. Regresi *Ridge* menjadi teknik untuk menganalisis data regresi berganda yang memiliki masalah multikolinearitas. Dengan menambahkan tingkat bias pada estimasi regresi, *Ridge* mengurangi kesalahan standar dan mendapatkan estimasi koefisien yang lebih akurat. Parameter regresi *Ridge* diestimasi dengan meminimalkan jumlah kuadrat kesalahan (*sum of square error*) yang menambahkan batasan pada kuadrat dengan mengecilkan koefisien mendekati nol. Regresi *Ridge* menghukum fitur-fitur yang memiliki kemiringan (*slope*) yang lebih tinggi yaitu dengan menambahkan penalti pada kuadrat dari besarnya koefisien (Enwere et al., 2023). Regresi *Ridge* dapat didefinisikan sebagai berikut:

$$\varepsilon^2 + \lambda \sum (\delta_1^2 + \delta_2^2 + \dots + \delta_n^2) \quad \lambda > 0 \quad (2.3)$$

Model regresi berganda dapat di regularisasi dengan menambahkan penalti regresi *Ridge* (L2 norm) sebagai berikut:

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \dots + \delta_n w_n + \varepsilon + \lambda \sum (\delta_1^2 + \delta_2^2 + \dots + \delta_n^2) \quad (2.4)$$



Dengan SSE dinyatakan sebagai berikut:

$$\varepsilon^2 = (v - \hat{v})^2 + \lambda \sum (\delta_1^2 + \delta_2^2 + \dots + \delta_n^2) \quad (2.5)$$

## 2.6. LASSO Regression

Pada tahun 1996 Robert Tibshirani mengusulkan sebuah metode baru untuk estimasi dalam model linier, dikarenakan regresi *Ridge* hanya menyusutkan koefisien mendekati nol tapi tidak menetapkan koefisien menjadi nol. *LASSO* (*Least Absolute Shrinkage and Selection Operator*) muncul sebagai metode baru yang dapat mengecilkan beberapa koefisien dan menetapkan yang lain ke nol. Dikarenakan hal ini *LASSO Regression* dapat melakukan *feature selection* dengan mempertahankan fitur korelasi tinggi dengan koefisien yang lebih tinggi tetapi tidak dengan koefisien yang lebih rendah (Enwere et al., 2023). Regresi *LASSO* dapat didefinisikan sebagai berikut:

$$\varepsilon^2 + \lambda \sum |\delta_1| + |\delta_2| + \dots + |\delta_n| \quad \lambda > 0 \quad (2.6)$$

Model regresi berganda dapat diregularisasikan dengan menambahkan penalti regresi *LASSO* (*L1 norm*) sebagai berikut:

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \dots + \delta_n w_n + \varepsilon + \lambda \sum |\delta_1| + |\delta_2| + \dots + |\delta_n| \quad (2.7)$$

Dengan SSE dinyatakan sebagai berikut:

$$\varepsilon^2 = (v - \hat{v})^2 + \lambda \sum |\delta_1| + |\delta_2| + \dots + |\delta_n| \quad (2.8)$$

## **BAB III**

### **ANALISIS DAN PERANCANGAN SISTEM**

#### **3.1. Analisis Sistem**

Mengacu pada langkah-langkah melakukan penelitian pada suatu sistem untuk mengidentifikasi kebutuhan komponennya. Fase analisis sistem mencakup pemeriksaan terhadap masalah untuk memastikan sistem. Selanjutnya, sepanjang tahap desain sistem, dilakukan analisis untuk mengidentifikasi proses dan sumber data yang dibutuhkan oleh perangkat lunak.

##### **3.1.1. Analisis Masalah**

Pemeriksaan sistem mengacu pada tindakan melakukan penelitian pada suatu sistem untuk mengidentifikasi kebutuhan komponennya. Ini adalah langkah penting dalam proses desain sistem secara keseluruhan. Fase analisis sistem mencakup pemeriksaan terhadap masalah untuk memastikan sistem. Selanjutnya, sepanjang tahap desain sistem, dilakukan analisis untuk mengidentifikasi proses dan sumber data yang dibutuhkan oleh sistem. Tahap pertama dalam menyelesaikan suatu masalah adalah dengan mengenalinya, sedangkan langkah berikutnya adalah memastikan permasalahan-permasalahannya. Berikut adalah beberapa kekhawatiran yang terdeteksi.

- a. Obesitas dan lemak tubuh berlebih dapat memicu beberapa penyakit kronis
- b. Metode pengukuran lemak tubuh secara tradisional cenderung tidak mudah diakses, invasif, dan tidak efektif.

##### **3.1.2. Analisis Requirements**

Pemeriksaan untuk mengidentifikasi apa saja yang akan diperlukan dalam merancang sistem yang diinginkan. Terdapat dua jenis analisis *requirements* yang dapat diidentifikasi dalam perancangan sistem yakni, *requirement* fungsional dan *requirement* non-fungsional.

### 1. Analisis *Requirement* Fungsional

Tahap ini berkaitan dengan identifikasi fungsionalitas yang diperlukan oleh sistem atau proyek yang sedang dikerjakan. Ini mencakup pemahaman terhadap apa yang harus dilakukan oleh sistem atau proyek, serta kebutuhan-kebutuhan fungsional yang harus dipenuhi untuk mencapai tujuan yang diinginkan.

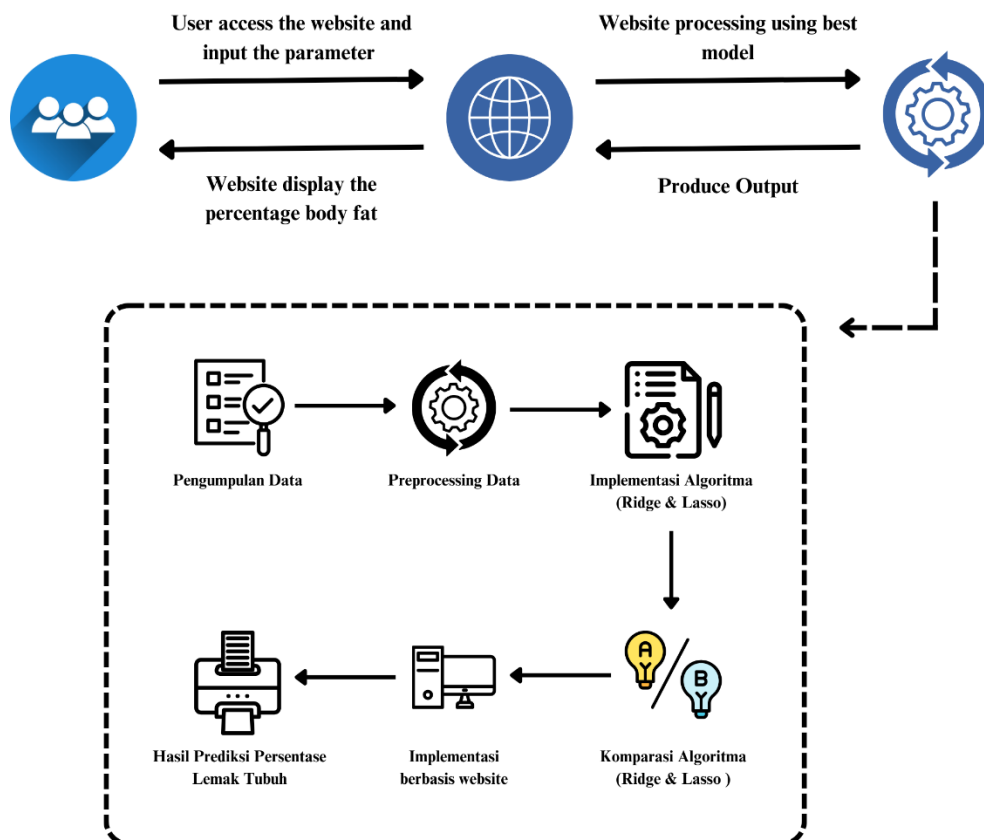
- a. Sistem memiliki kemampuan untuk memprediksi persentase lemak tubuh dengan menggunakan inputan yang dilakukan oleh *user* melalui algoritma terbaik.
- b. Sistem mampu menampilkan hasil dari prediksi persentase lemak tubuh sesuai dengan inputan.

### 2. Analisis *Requirement* Non-Fungsional

Tahap ini berkaitan dengan identifikasi non-fungsional yang harus dipenuhi oleh sistem atau proyek. Ini meliputi hal-hal yang tidak berkaitan langsung dengan fungsionalitas sistem, tetapi tetap penting untuk keberhasilan dan kepuasan pengguna sistem.

- a. Sistem yang telah dikembangkan memiliki *interface* yang mudah digunakan oleh pengguna.
- b. Sistem hanya menghasilkan output yang sesuai apabila semua inputan diisi tanpa terkecuali.
- c. Output dari sistem hanya menampilkan persentase lemak tubuh.

### 3.1.3. Arsitektur Sistem



**Gambar 3.1** Arsitektur Sistem

Arsitektur umum sistem ini secara umum bekerja sebagai berikut:

1. *User* mengakses *website* dan memasukkan inputan yang diminta.
2. Kemudian sistem memproses inputan tersebut menggunakan algoritma terbaik.
3. Ketika sistem telah selesai menghitung prediksi persentase lemak tubuh menggunakan algoritma terbaik maka menghasilkan output.
4. Output yang ditampilkan pada *website* berupa prediksi persentase lemak tubuh.

Adapun proses yang terjadi pada pemilihan algoritma terbaik adalah sebagai berikut:

1. Pengumpulan data dimana dataset didapatkan dari *website kaggle* bernama *bodyfat prediction*, yang memuat berbagai pengukuran lingkaran tubuh pada 252 pria.

2. *Preprocessing* data adalah tahap dimana dilakukan cek dan pembersihan data, meliputi cek *missing values*, *duplicated* data, dan *outliers*. Dilakukan juga konversi satuan nilai dan perhitungan BMI.
3. Algoritma *Ridge* dan *LASSO Regression* diimplementasikan untuk membuat model yang dapat melakukan prediksi berdasarkan data latih, *best alpha* dicari menggunakan metode *grid search*.
4. Algoritma *Ridge* dan *LASSO Regression* selanjutnya melakukan prediksi pada data uji, dan dilakukan evaluasi dengan membandingkan MSE, MAE,  $R^2$ , dan *Adjusted R<sup>2</sup>*, lalu dipilih satu algoritma terbaik diantara *Ridge* dan *LASSO Regression*.
5. Algoritma terbaik tersebut lalu diimplementasikan pada *website*, dengan bahasa pemrograman HTML dan CSS.
6. Pada *website* tersebut, *user* dapat menginputkan variabel yang nantinya diproses menggunakan algoritma terbaik dan menghasilkan output berupa prediksi persentase lemak tubuh.

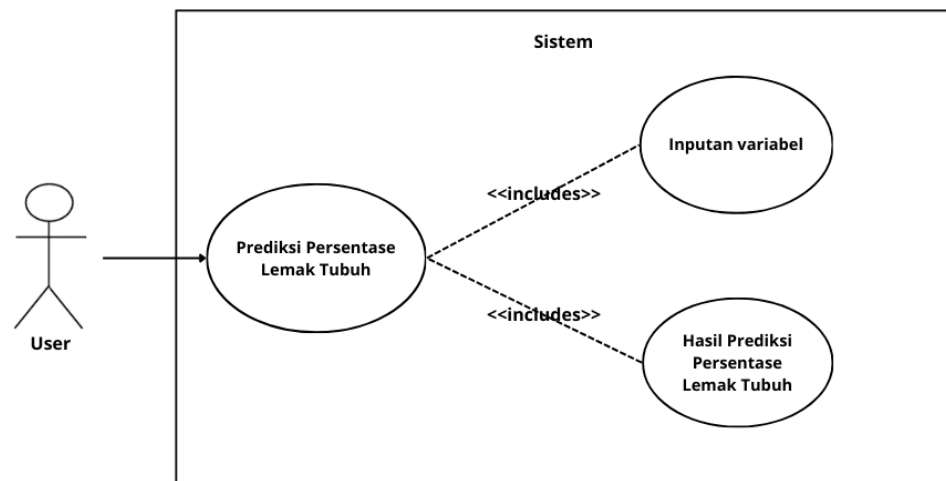
### 3.2. Pemodelan Sistem

Pemodelan sistem merupakan cara menentukan fungsi serta tujuan utama suatu sistem supaya dapat mengetahui kebutuhan user dan gambaran umum dari sebuah sistem.

#### 3.2.1. Use Case Diagram

Use Case Diagram merupakan gambaran bagi pengguna untuk mengetahui bagaimana cara menggunakan sistem.



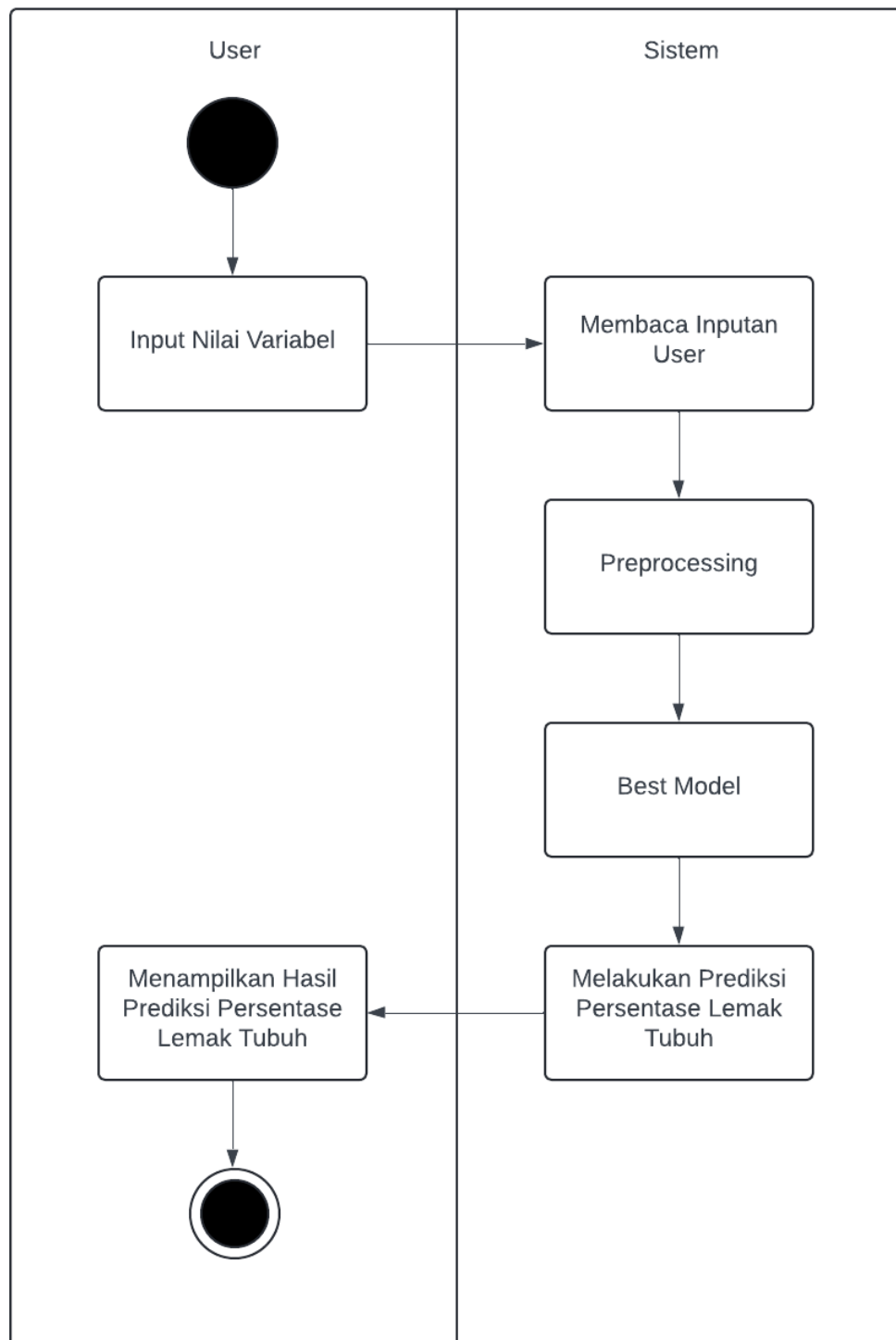


**Gambar 3.2** Use Case Diagram

Pada gambar 3.2 terlihat bahwa skenario penggunaan sistem ini dimulai dengan langkah pertama dimana pengguna membuka situs web sistem. Setelah itu, *user* diminta untuk mengisi inputan variabel yang dikelola sistem untuk melakukan prediksi. Selanjutnya, sistem menjalankan proses prediksi dengan menggunakan model algoritma terbaik. Terakhir, sistem menampilkan hasil prediksi persentase lemak tubuh kepada *user*.

### 3.2.2. Activity Diagram

*Activity* diagram adalah representasi aktivitas sistem yang menunjukkan langkah-langkah yang terlibat dalam operasi sistem dari awal hingga akhir.



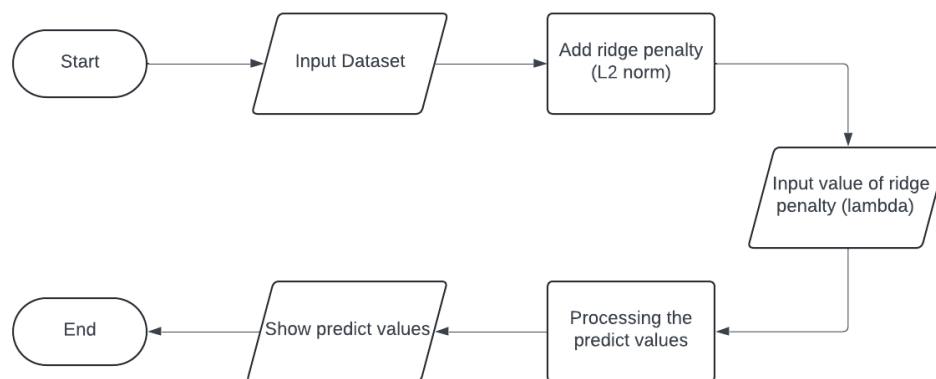
**Gambar 3.3** Activity Diagram

Gambar 3.3 menunjukkan bahwa langkah awal penggunaan sistem ini dimulai ketika *user* menginputkan variabel yang diminta pada *form*. Setelah semua inputan diisi, selanjutnya *user* perlu menekan tombol “*Prediction*” agar sistem dapat menjalankan proses prediksi dan menampilkan hasil prediksi persentase lemak tubuh yang sesuai.

### 3.3. Flowchart

*Flowchart* atau diagram alir adalah representasi visual dari prosedur yang terlibat dalam menemukan solusi untuk suatu masalah. Setiap prosedur ditampilkan sebagai diagram dengan garis dan anak panah yang menghubungkan berbagai langkah. Dengan menggunakan *flowchart*, masalah dapat dipecahkan langkah demi langkah hingga hasil yang diinginkan tercapai.

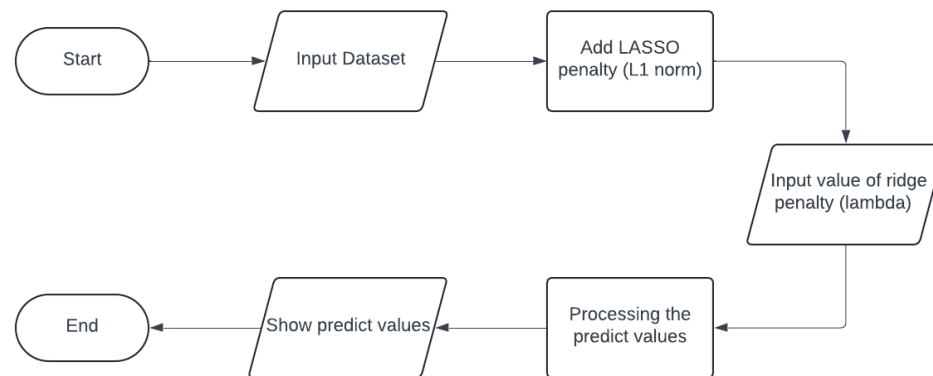
#### 3.3.1. Flowchart Ridge Regression



**Gambar 3.4** Flowchart Ridge Regression

Berdasarkan *flowchart* pada gambar 3.4, cara kerja dari *Ridge Regression* yaitu dengan menambahkan nilai penalti (*L2 norm*), dalam penelitian ini peneliti mencari nilai penalti terbaik (*best lambda*) menggunakan *grid search*. Setelah mendapatkan *best lambda*, nilai tersebut lalu digunakan dalam model pelatihan yang nantinya menghasilkan output prediksi persentase lemak tubuh.

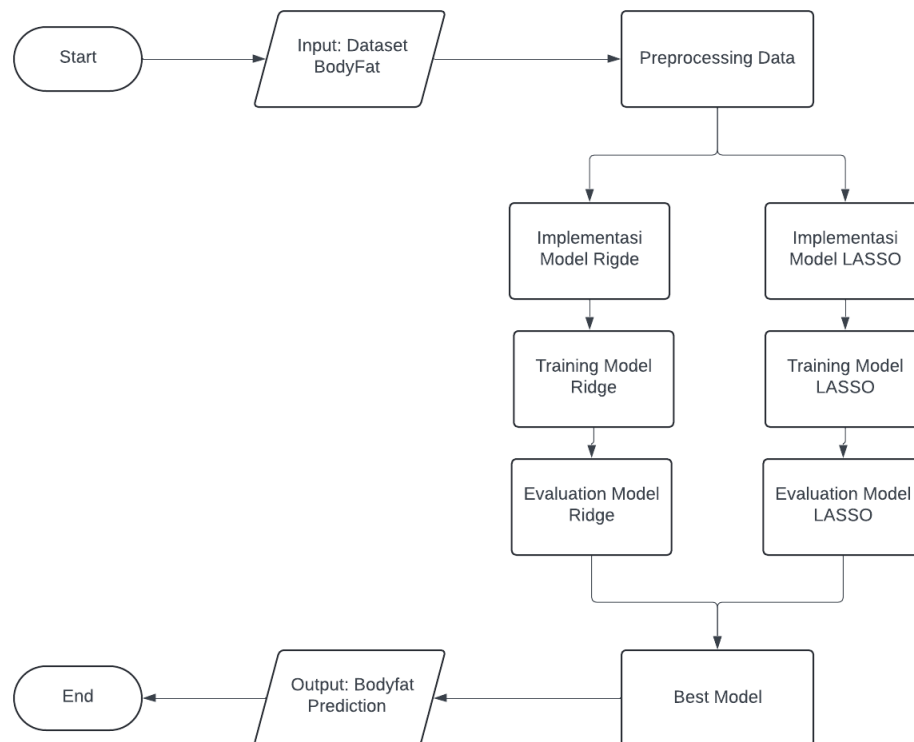
### 3.3.2. Flowchart LASSO Regression



**Gambar 3.5** Flowchart LASSO Regression

Berdasarkan *flowchart* pada gambar 3.5, cara kerja dari *LASSO Regression* yaitu dengan menambahkan nilai penalti (*L1 norm*), dalam penelitian ini peneliti mencari nilai penalti terbaik (*best lambda*) menggunakan *grid search*. Setelah mendapatkan *best lambda*, nilai tersebut lalu digunakan dalam model pelatihan yang nantinya menghasilkan output prediksi persentase lemak tubuh.

### 3.3.3. Flowchart Sistem



**Gambar 3.6** Flowchart Sistem

Berdasarkan *flowchart* pada gambar 3.6, cara kerja dari sistem yaitu pertama membaca dataset *bodyfat*, lalu melakukan *preprocessing* data. Setelah itu data akan diimplementasikan pada tiap-tiap algoritma yaitu *Ridge* dan *LASSO Regression*, melalui training model dan dilakukan evaluasi. Lalu dipilih algoritma terbaik (*best model*) berdasarkan perbandingan hasil evaluasi sehingga menghasilkan output prediksi persentase lemak tubuh.

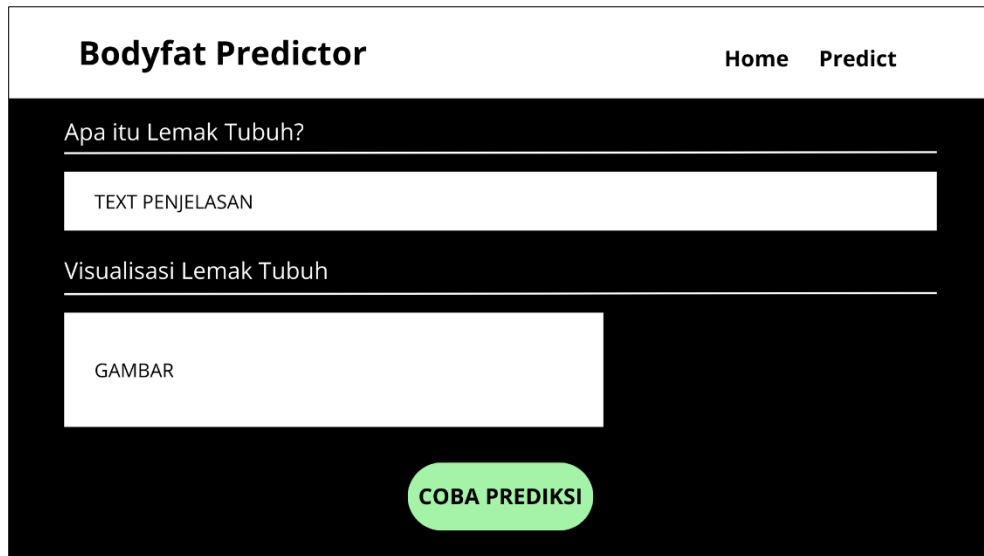
### 3.4. Perancangan Interface

Perancangan *interface* merupakan suatu cara yang penting untuk diperhatikan. *Interface* ialah tampilan antarmuka yang menghubungkan *user* ke dalam sistem. *Interface* harus menarik agar *user* tertarik untuk menggunakannya. *Interface* harus juga mudah dimengerti agar *user* mudah menggunakan aplikasi. Pada penelitian ini sistem dibuat berbasis *website* agar mudah di akses dimana saja. Sistem ini memiliki halaman untuk menerima inputan variabel dalam bentuk *form* yang digunakan untuk melakukan prediksi persentase lemak tubuh.



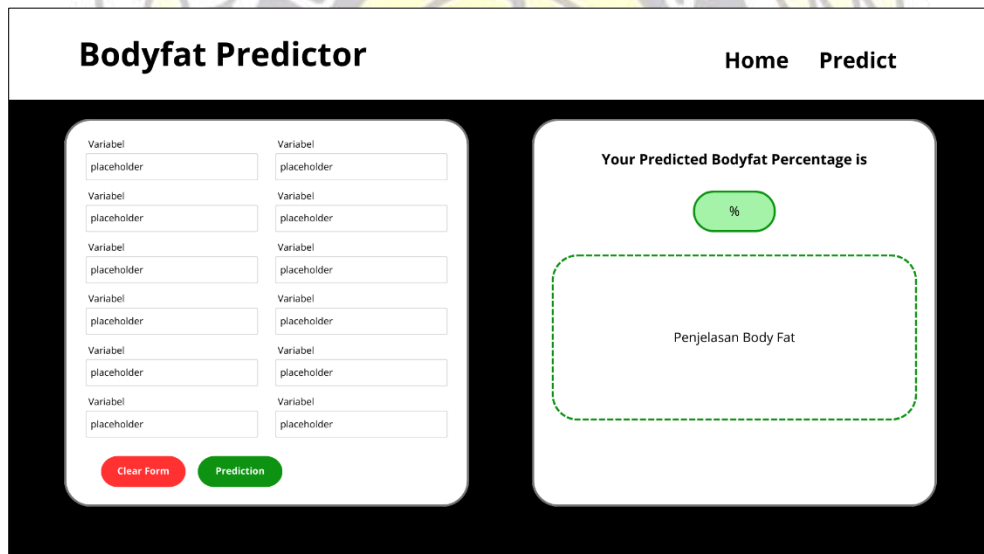
### 3.4.1. Interface Halaman Tampilan Sistem

Gambar 3.7 dan 3.8 ini merupakan tampilan dari sistem yang dibuat.



The screenshot shows the 'Bodyfat Predictor' application interface on the 'Home' page. At the top, there is a navigation bar with the title 'Bodyfat Predictor' and two links: 'Home' and 'Predict'. Below the navigation bar, the main content area has a dark background. It starts with the text 'Apa itu Lemak Tubuh?' followed by a white rectangular box containing the text 'TEXT PENJELASAN'. Below this is another section titled 'Visualisasi Lemak Tubuh' with a white rectangular box containing the text 'GAMBAR'. At the bottom center, there is a prominent green button with the text 'COBA PREDIKSI' in white capital letters.

**Gambar 3.7** Interface Halaman Home



The screenshot shows the 'Bodyfat Predictor' application interface on the 'Predict' page. The navigation bar at the top is identical to the home page, with 'Bodyfat Predictor' and links to 'Home' and 'Predict'. The main content area has a dark background and is divided into two main sections. On the left, there is a white rounded rectangle containing a form with 16 input fields, each labeled 'Variabel' above it. At the bottom of this form are two buttons: a red 'Clear Form' button and a green 'Prediction' button. On the right, there is another white rounded rectangle. At the top of this section, it says 'Your Predicted Bodyfat Percentage is' followed by a green circle containing a '%' symbol. Below this is a large dashed green rectangular box containing the text 'Penjelasan Body Fat'.

**Gambar 3.8** Interface Halaman Predict

Keterangan Gambar:

1. *Navigation Bar*, Memudahkan pindah halaman antara halaman *home* dan *predict*.
2. *Button*, Coba Prediksi, untuk pindah ke halaman predict.

3. *Form*, Menampung inputan variabel.
4. *Button, Prediction* untuk melakukan proses prediksi.
5. *Button, Clear Form* untuk membersihkan nilai pada *form*.
6. *Result Box*, Menampilkan hasil prediksi persentase lemak tubuh.
7. *Text Box*, Penjelasan persentase lemak tubuh ideal.



## BAB IV

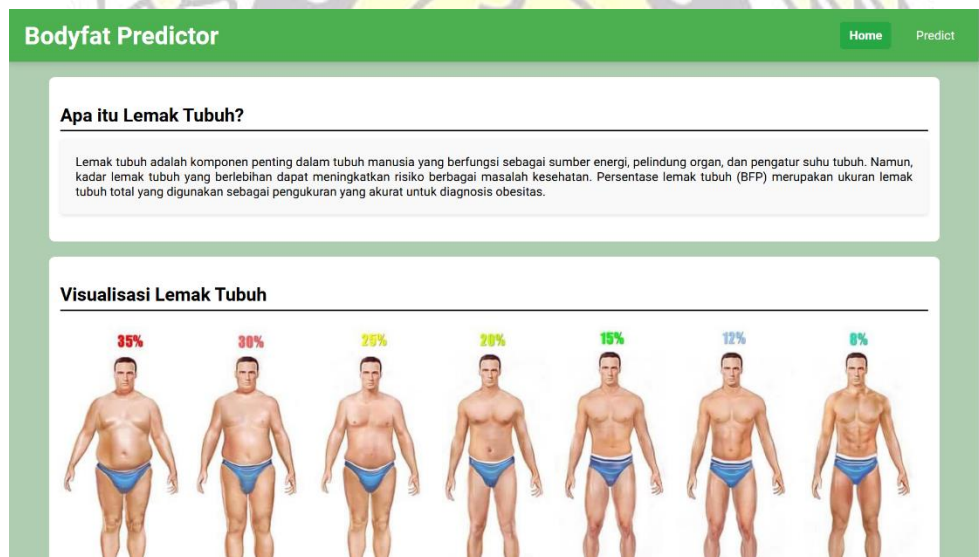
### IMPLEMENTASI DAN PENGUJIAN SISTEM

#### 4.1. Implementasi Sistem

Hasil dari penelitian ini adalah sistem yang dikembangkan menggunakan bahasa pemrograman *python*. Dalam sistem ini proses implementasi terjadi pada halaman utama dimana *user* dapat menginput nilai untuk diolah melalui proses prediksi yang nantinya menampilkan hasil prediksi.

##### 4.1.1. Halaman Tampilan Sistem

Halaman sistem ini menampilkan halaman awal yang muncul kepada *user* ketika menjalankan program.



**Gambar 4.1** Halaman Home

**Bodyfat Predictor** Home Predict

Usia  
Enter your Age

Berat Badan  
Enter in kg

Tinggi Badan  
Enter in cm

BMI  
BMI will be calculated automatically

Lingkar Leher  
Enter in cm

Lingkar Dada  
Enter in cm

Lingkar Perut  
Enter in cm

Lingkar Pinggul  
Enter in cm

Lingkar Paha  
Enter in cm

Lingkar Lutut  
Enter in cm

Lingkar Pergelangan Kaki  
Enter in cm

Lingkar Bisep  
Enter in cm

Lingkar Lengan Bawah  
Enter in cm

Lingkar Pergelangan Tangan  
Enter in cm

Clear Form Prediction

**Your Predicted Bodyfat Percentage is**

0%

Body fat percentage is a key indicator of good health. A high body fat percentage may put you at a higher risk of lifestyle diseases. Best body fat percentages averaged between 12% and 20% for men and 20% and 30% for women.  
*Dikutip dari The American Journal of Clinical Nutrition berjudul "Healthy body weights: an alternative perspective"*

| Kategori | Pria (%) | Wanita (%) |
|----------|----------|------------|
| Atlet    | 6-13     | 14-20      |
| Bugar    | 14-17    | 21-14      |
| Normal   | 18-24    | 25-31      |
| Obese    | >25      | >32        |

**Gambar 4.2** Halaman Predict

#### 4.1.2. Halaman Tampilan Hasil Prediksi

Halaman ini menampilkan halaman yang muncul kepada *user* ketika sudah menginputkan variabel dan menekan tombol “*prediction*” dan menampilkan hasil dari prediksi persentase lemak tubuh.

**Bodyfat Predictor** Home Predict

Usia  
24.0

Berat Badan  
83.57

Tinggi Badan  
180.98

BMI  
25.51

Lingkar Leher  
34.4

Lingkar Dada  
97.3

Lingkar Perut  
100.0

Lingkar Pinggul  
101.0

Lingkar Paha  
63.2

Lingkar Lutut  
42.2

Lingkar Pergelangan Kaki  
24.0

Lingkar Bisep  
32.2

Lingkar Lengan Bawah  
27.7

Lingkar Pergelangan Tangan  
17.7

**Your Predicted Bodyfat Percentage is**

24.85%

Body fat percentage is a key indicator of good health. A high body fat percentage may put you at a higher risk of lifestyle diseases. Best body fat percentages averaged between 12% and 20% for men and 20% and 30% for women.  
*Dikutip dari The American Journal of Clinical Nutrition berjudul "Healthy body weights: an alternative perspective"*

| Kategori | Pria (%) | Wanita (%) |
|----------|----------|------------|
| Atlet    | 6-13     | 14-20      |
| Bugar    | 14-17    | 21-14      |
| Normal   | 18-24    | 25-31      |
| Obese    | >25      | >32        |

**Gambar 4.3** Tampilan Hasil Prediksi

#### 4.2. Pra-Pemrosesan Dataset

Berikut adalah penjelasan dari *preprocessing* dataset dalam sistem. Pertama, membaca dataset *bodyfat* dalam file csv seperti pada gambar 4.3 ini.



```
1 df_raw = pd.read_csv('bodyfat.csv')
2 df_raw.head()
```

|   | Density | BodyFat | Age | Weight | Height | Neck | Chest | Abdomen | Hip   | Thigh | Knee | Ankle | Biceps | Forearm | Wrist |
|---|---------|---------|-----|--------|--------|------|-------|---------|-------|-------|------|-------|--------|---------|-------|
| 0 | 1.0708  | 12.3    | 23  | 154.25 | 67.75  | 36.2 | 93.1  | 85.2    | 94.5  | 59.0  | 37.3 | 21.9  | 32.0   | 27.4    | 17.1  |
| 1 | 1.0853  | 6.1     | 22  | 173.25 | 72.25  | 38.5 | 93.6  | 83.0    | 98.7  | 58.7  | 37.3 | 23.4  | 30.5   | 28.9    | 18.2  |
| 2 | 1.0414  | 25.3    | 22  | 154.00 | 66.25  | 34.0 | 95.8  | 87.9    | 99.2  | 59.6  | 38.9 | 24.0  | 28.8   | 25.2    | 16.6  |
| 3 | 1.0751  | 10.4    | 26  | 184.75 | 72.25  | 37.4 | 101.8 | 86.4    | 101.2 | 60.1  | 37.3 | 22.8  | 32.4   | 29.4    | 18.2  |
| 4 | 1.0340  | 28.7    | 24  | 184.25 | 71.25  | 34.4 | 97.3  | 100.0   | 101.9 | 63.2  | 42.2 | 24.0  | 32.2   | 27.7    | 17.7  |

**Gambar 4.4** Import Dataset

Setelah itu, dilakukan pengecekan *missing values* dan juga *duplicated* data untuk memastikan tidak ada data yang hilang atau kosong sekaligus tidak ada data yang bernilai sama. Seperti pada gambar 4.4 dan gambar 4.5 ini.

```
1 # Menghitung missing values
2 missing_values = df_raw.isnull().sum()
3 transposed_missing_values = missing_values.to_frame(name='Missing Values').T
4 transposed_missing_values
```

|                | Density | BodyFat | Age | Weight | Height | Neck | Chest | Abdomen | Hip | Thigh | Knee | Ankle | Biceps | Forearm | Wrist |
|----------------|---------|---------|-----|--------|--------|------|-------|---------|-----|-------|------|-------|--------|---------|-------|
| Missing Values | 0       | 0       | 0   | 0      | 0      | 0    | 0     | 0       | 0   | 0     | 0    | 0     | 0      | 0       | 0     |

**Gambar 4.5** Cek Missing Values

```
1 # Melakukan pengecekan duplikasi data
2 # false berarti tidak ada duplikat data
3 df_raw.duplicated().values.any()
```

False

**Gambar 4.6** Cek Duplicated Data

Setelah memastikan tidak ada *missing values* dan *duplicated* data, dilakukan juga pengecekan tipe data dan *describe* data, untuk melihat statistik dasar dari data. Pada saat mengecek *describe* data, penulis melihat data anomali yaitu nilai minimum dari *bodyfat* adalah 0. Oleh dikarenakan hal tersebut, penulis juga melakukan cek *outliers* menggunakan *z-score* dan ditemukan terdapat satu *outliers*. Penulis memutuskan untuk menghapus kedua baris data tersebut, sehingga jumlah total baris adalah 250. Seperti pada gambar 4.6 dan gambar 4.7 ini

```

1 # melihat nilai bodyfat = 0
2 # untungnya cuman ada 1 baris
3 zero_bodyfat_rows = df_raw[df_raw['BodyFat'] == 0]
4 zero_bodyfat_rows

```

|     | Density | BodyFat | Age | Weight | Height | Neck | Chest | Abdomen | Hip  | Thigh | Knee | Ankle | Biceps | Forearm | Wrist |
|-----|---------|---------|-----|--------|--------|------|-------|---------|------|-------|------|-------|--------|---------|-------|
| 181 | 1.1089  | 0.0     | 40  | 118.5  | 68.0   | 33.8 | 79.3  | 69.4    | 85.0 | 47.2  | 33.5 | 20.2  | 27.7   | 24.6    | 16.5  |

**Gambar 4.7** Anomali Data Bodyfat bernilai 0

```

1 from scipy import stats
2 z_scores = stats.zscore(df_raw['BodyFat'])
3 df_raw['z_score'] = z_scores
4
5 outliers = df_raw[df_raw['z_score'].abs() > 3]
6 outliers

```

|     | Density | BodyFat | Age | Weight | Height | Neck | Chest | Abdomen | Hip   | Thigh | Knee | Ankle | Biceps | Forearm | Wrist | z_score  |
|-----|---------|---------|-----|--------|--------|------|-------|---------|-------|-------|------|-------|--------|---------|-------|----------|
| 215 | 0.995   | 47.5    | 51  | 219.0  | 64.0   | 41.2 | 119.8 | 122.1   | 112.8 | 62.5  | 36.9 | 23.6  | 34.7   | 29.1    | 18.4  | 3.394253 |

**Gambar 4.8** Outlier

Kemudian, penulis juga melakukan konversi satuan untuk variabel *weight* dan *height*, yang awalnya dari lbs ke kg, dan dari inch ke cm. Seperti gambar 4.8 dan gambar 4.9 ini.

```

1 # Ubah satuan weight (lbs -> kg)
2 # kg = lbs × 0.453592
3 # disimpan ke kolom bernama "Weight_kg"
4 df_raw['Weight_kg'] = (df_raw['Weight'] * 0.453592).round(2)

```

**Gambar 4.9** Konversi Satuan Weight

```

1 # Ubah satuan height (inch -> cm)
2 # cm = inch × 2.54
3 # disimpan ke kolom bernama "Height_cm"
4 df_raw['Height_cm'] = (df_raw['Height'] * 2.54).round(2)

```

**Gambar 4.10** Konversi Satuan Height

Kemudian, penulis melakukan perhitungan BMI seperti pada gambar 4.10 ini.

```

1 # BMI = berat_badan_kg / (tinggi_m) 2
2 # Mengonversi tinggi dari cm ke meter
3 df_raw['Height_m'] = df_raw['Height_cm'] / 100
4
5 # Menghitung nilai BMI
6 df_raw['BMI'] = (df_raw['Weight_kg'] / (df_raw['Height_m'] ** 2)).round(2)

```

**Gambar 4.11** Perhitungan BMI

Setelah melakukan preparation data yang dibutuhkan, penulis kemudian menyimpan variabel yang dibutuhkan untuk *modelling* bernama “df” dengan total 250 baris dan 15 kolom termasuk kolom *bodyfat*. Seperti pada gambar 4.11 ini.

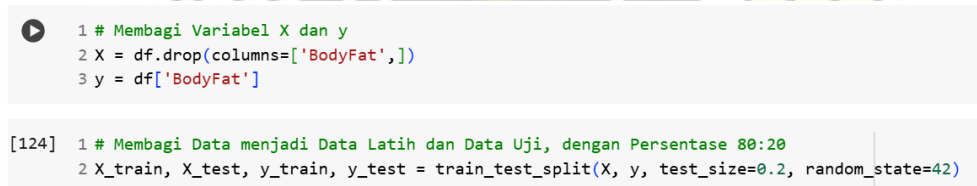


```
1 df.head()
```

|   | BodyFat | Age | Weight_kg | Height_cm | BMI   | Neck | Chest | Abdomen | Hip   | Thigh | Knee | Ankle | Biceps | Forearm | Wrist |
|---|---------|-----|-----------|-----------|-------|------|-------|---------|-------|-------|------|-------|--------|---------|-------|
| 0 | 12.3    | 23  | 69.97     | 172.08    | 23.63 | 36.2 | 93.1  | 85.2    | 94.5  | 59.0  | 37.3 | 21.9  | 32.0   | 27.4    | 17.1  |
| 1 | 6.1     | 22  | 78.58     | 183.52    | 23.33 | 38.5 | 93.6  | 83.0    | 98.7  | 58.7  | 37.3 | 23.4  | 30.5   | 28.9    | 18.2  |
| 2 | 25.3    | 22  | 69.85     | 168.28    | 24.67 | 34.0 | 95.8  | 87.9    | 99.2  | 59.6  | 38.9 | 24.0  | 28.8   | 25.2    | 16.6  |
| 3 | 10.4    | 26  | 83.80     | 183.52    | 24.88 | 37.4 | 101.8 | 86.4    | 101.2 | 60.1  | 37.3 | 22.8  | 32.4   | 29.4    | 18.2  |
| 4 | 28.7    | 24  | 83.57     | 180.98    | 25.51 | 34.4 | 97.3  | 100.0   | 101.9 | 63.2  | 42.2 | 24.0  | 32.2   | 27.7    | 17.7  |

**Gambar 4.12** Dataset Model

Lalu penulis melakukan *train test split* atau pembagian data latih dan data uji dengan persentase 80:20, dimana kolom *bodyfat* sebagai target (y), dan sisa kolom lainnya sebagai fitur (X). Seperti pada gambar 4.12 ini.

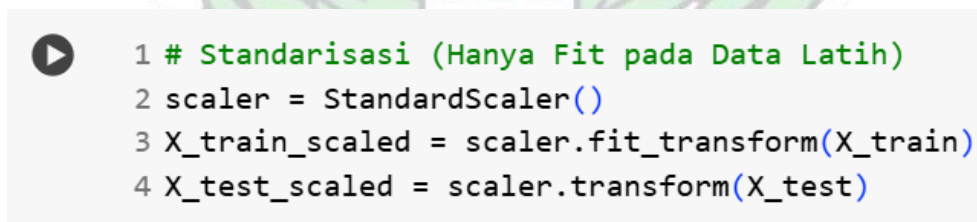


```
1 # Membagi Variabel X dan y
2 X = df.drop(columns=['BodyFat'],)
3 y = df['BodyFat']

[124] 1 # Membagi Data menjadi Data Latih dan Data Uji, dengan Persentase 80:20
      2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

**Gambar 4.13** Train Test Split

Setelah data dibagi menjadi data latih dan data uji, penulis juga melakukan standarisasi untuk data pada variabel X (fitur). Seperti pada gambar 4.13 ini.



```
1 # Standarisasi (Hanya Fit pada Data Latih)
2 scaler = StandardScaler()
3 X_train_scaled = scaler.fit_transform(X_train)
4 X_test_scaled = scaler.transform(X_test)
```

**Gambar 4.14** Standarisasi Data

### 4.3. Implementasi Algoritma Ridge dan LASSO Regression

Sebelum melakukan pemodelan dan pelatihan menggunakan data latih, penulis terlebih dahulu mencari nilai penalti ( $\lambda$ /alpha) terbaik untuk kedua algoritma menggunakan *grid search*. Seperti pada gambar 4.14 ini.

```

1 # Membuat rentang nilai alpha yang akan diuji
2 alpha_values = {'alpha': [0.001, 0.01, 0.1, 1, 10, 100, 1000]}
3
4 # Ridge Regression dengan Grid Search
5 ridge_model = Ridge()
6 grid_search_ridge = GridSearchCV(ridge_model, alpha_values, cv=5, scoring='neg_mean_squared_error')
7 grid_search_ridge.fit(X_train_scaled, y_train)
8
9 # Lasso Regression dengan Grid Search
10 lasso_model = Lasso()
11 grid_search_lasso = GridSearchCV(lasso_model, alpha_values, cv=5, scoring='neg_mean_squared_error')
12 grid_search_lasso.fit(X_train_scaled, y_train)
13
14 # Menampilkan nilai alpha terbaik untuk masing-masing model
15 best_alpha_ridge = grid_search_ridge.best_params_['alpha']
16 best_alpha_lasso = grid_search_lasso.best_params_['alpha']
17 print(f"Alpha terbaik untuk Ridge Regression: {best_alpha_ridge}")
18 print(f"Alpha terbaik untuk Lasso Regression: {best_alpha_lasso}")

```

Alpha terbaik untuk Ridge Regression: 10  
Alpha terbaik untuk Lasso Regression: 0.1

Gambar 4.15 Grid Search

Setelah nilai alpha terbaik didapatkan menggunakan *grid search*, barulah dilakukan pelatihan model menggunakan data latih. Seperti gambar 4.15 ini.

```

1 # Melatih model Ridge Regression dengan alpha terbaik
2 ridge_best_model = Ridge(alpha=best_alpha_ridge)
3 ridge_best_model.fit(X_train_scaled, y_train)

```

Ridge (i ?)  
Ridge(alpha=10)

```

[128] 1 # Melatih model Lasso Regression dengan alpha terbaik
      2 lasso_best_model = Lasso(alpha=best_alpha_lasso)
      3 lasso_best_model.fit(X_train_scaled, y_train)

```

Lasso (i ?)  
Lasso(alpha=0.1)

Gambar 4.16 Train Model Ridge dan LASSO



Kedua algoritma ini baik *Ridge* dan *LASSO Regression* bekerja dengan meminimalkan nilai fitur, namun terdapat perbedaan yang signifikan diantara keduanya. *Ridge Regression* hanya meminimalkan nilai fitur mendekati nol, berbeda dengan *LASSO Regression* yang dapat meminimalkan nilai fitur hingga nol. Berikut adalah nilai fitur dari tiap-tiap algoritma. Nilai koefisien untuk tiap fitur pada masing-masing algoritma dapat dilihat pada tabel 4.1 ini.

**Tabel 4.1** Nilai Koefisien

| Variabel X (Fitur) | Ridge  | LASSO  |
|--------------------|--------|--------|
| Age                | 1.317  | 0.853  |
| Weight             | 0.122  | -0     |
| Height             | -1.437 | -0.987 |
| BMI                | -0.877 | -0.380 |
| Neck               | -0.496 | -0.372 |
| Chest              | 0.332  | -0.029 |
| Abdomen            | 6.235  | 7.715  |
| Hip                | -0.305 | -0.370 |
| Thigh              | 0.922  | 0.196  |
| Knee               | 0.417  | 0      |
| Ankle              | 0.074  | 0      |
| Bicep              | 0.539  | 0.437  |
| Forearm            | 0.417  | 0.269  |
| Wrist              | -1.796 | -1.668 |

*Ridge* dan *LASSO regression* bekerja dengan menambahkan penalti pada nilai koefisien, namun terdapat perbedaan dimana *Ridge Regression* hanya meminimalkannya sedangkan *LASSO Regression* dapat membuatnya menjadi nol. Berikut adalah penjelasan bagaimana *Ridge* dan *LASSO Regression* bekerja dengan mengambil 25 baris pertama dari dataset sebagai sampel.

| BodyFat | Age | Weight | Height | BMI   | Neck | Chest | Abdomen | Hip   | Thigh | Knee | Ankle | Biceps | Forearm | Wrist |
|---------|-----|--------|--------|-------|------|-------|---------|-------|-------|------|-------|--------|---------|-------|
| 12,3    | 23  | 69,97  | 172,08 | 23,63 | 36,2 | 93,1  | 85,2    | 94,5  | 59    | 37,3 | 21,9  | 32     | 27,4    | 17,1  |
| 6,1     | 22  | 78,58  | 183,52 | 23,33 | 38,5 | 93,6  | 83      | 98,7  | 58,7  | 37,3 | 23,4  | 30,5   | 28,9    | 18,2  |
| 25,3    | 22  | 69,85  | 168,28 | 24,67 | 34   | 95,8  | 87,9    | 99,2  | 59,6  | 38,9 | 24    | 28,8   | 25,2    | 16,6  |
| 10,4    | 26  | 83,8   | 183,52 | 24,88 | 37,4 | 101,8 | 86,4    | 101,2 | 60,1  | 37,3 | 22,8  | 32,4   | 29,4    | 18,2  |
| 28,7    | 24  | 83,57  | 180,98 | 25,51 | 34,4 | 97,3  | 100     | 101,9 | 63,2  | 42,2 | 24    | 32,2   | 27,7    | 17,7  |
| 20,9    | 24  | 95,37  | 189,86 | 26,46 | 39   | 104,5 | 94,4    | 107,8 | 66    | 42   | 25,6  | 35,7   | 30,6    | 18,8  |
| 19,2    | 26  | 82,1   | 177,16 | 26,16 | 36,4 | 105,1 | 90,7    | 100,3 | 58,4  | 38,3 | 22,9  | 31,9   | 27,8    | 17,7  |
| 12,4    | 25  | 79,83  | 184,15 | 23,54 | 37,8 | 99,6  | 88,5    | 97,1  | 60    | 39,4 | 23,2  | 30,5   | 29      | 18,8  |
| 4,1     | 25  | 86,64  | 187,96 | 24,52 | 38,1 | 100,9 | 82,5    | 99,9  | 62,9  | 38,3 | 23,8  | 35,9   | 31,1    | 18,2  |
| 11,7    | 23  | 89,92  | 186,69 | 25,8  | 42,1 | 99,6  | 88,6    | 104,1 | 63,1  | 41,7 | 25    | 35,6   | 30      | 19,2  |
| 7,1     | 26  | 84,48  | 189,23 | 23,59 | 38,5 | 101,5 | 83,6    | 98,2  | 59,7  | 39,7 | 25,2  | 32,8   | 29,4    | 18,5  |
| 7,8     | 27  | 97,98  | 193,04 | 26,29 | 39,4 | 103,6 | 90,9    | 107,7 | 66,2  | 39,2 | 25,9  | 37,2   | 30,2    | 19    |
| 20,8    | 32  | 81,87  | 176,53 | 26,27 | 38,4 | 102   | 91,6    | 103,9 | 63,4  | 38,3 | 21,5  | 32,5   | 28,6    | 17,7  |
| 21,2    | 30  | 93,1   | 180,98 | 28,42 | 39,4 | 104,1 | 101,8   | 108,6 | 66    | 41,5 | 23,7  | 36,9   | 31,6    | 18,8  |
| 22,1    | 35  | 85,16  | 176,53 | 27,33 | 40,5 | 101,3 | 96,4    | 100,1 | 69    | 39   | 23,1  | 36,1   | 30,5    | 18,2  |
| 20,9    | 35  | 73,82  | 167,64 | 26,27 | 36,4 | 99,1  | 92,8    | 99,2  | 63,1  | 38,7 | 21,7  | 31,1   | 26,4    | 16,9  |
| 29      | 34  | 88,79  | 180,34 | 27,3  | 38,9 | 101,9 | 96,4    | 105,2 | 64,8  | 40,8 | 23,1  | 36,2   | 30,8    | 17,3  |
| 22,9    | 32  | 94,91  | 180,34 | 29,18 | 42,1 | 107,6 | 97,5    | 107   | 66,9  | 40   | 24,4  | 38,2   | 31,6    | 19,3  |
| 16      | 28  | 83,35  | 172,08 | 28,15 | 38   | 106,8 | 89,6    | 102,4 | 64,2  | 38,7 | 22,9  | 37,2   | 30,5    | 18,5  |
| 16,5    | 33  | 96,05  | 186,69 | 27,56 | 40   | 106,2 | 100,5   | 109   | 65,8  | 40,6 | 24    | 37,1   | 30,1    | 18,2  |
| 19,1    | 28  | 81,19  | 172,72 | 27,22 | 39,1 | 103,3 | 95,9    | 104,9 | 63,5  | 38   | 22,1  | 32,5   | 30,3    | 18,4  |
| 15,2    | 28  | 90,95  | 177,16 | 28,98 | 41,3 | 111,4 | 98,8    | 104,8 | 63,4  | 40,6 | 24,6  | 33     | 32,8    | 19,9  |
| 15,6    | 31  | 63,62  | 173,36 | 21,17 | 33,9 | 86    | 76,4    | 94,6  | 57,4  | 35,3 | 22,2  | 27,9   | 25,9    | 16,7  |
| 17,7    | 32  | 67,47  | 177,8  | 21,34 | 35,5 | 86,7  | 80      | 93,4  | 54,9  | 36,2 | 22,1  | 29,8   | 26,7    | 17,1  |
| 14      | 28  | 68,61  | 172,08 | 23,17 | 34,5 | 90,2  | 76,3    | 95,8  | 58,4  | 35,5 | 22,9  | 31,1   | 28      | 17,6  |

**Gambar 4.17 Sampel Data**

Lakukan standarisasi data terlebih dahulu hanya pada variabel X, sedangkan variabel Y (*Bodyfat*) tidak dilakukan standarisasi, data yang sudah distandarisasi dapat dilihat pada gambar 4.17 ini.

| Age   | Weight | Height | BMI   | Neck  | Chest | Abdomen | Hip   | Thigh | Knee  | Ankle | Biceps | Forearm | Wrist |
|-------|--------|--------|-------|-------|-------|---------|-------|-------|-------|-------|--------|---------|-------|
| -1,23 | -1,36  | -1,11  | -0,92 | -0,77 | -1,13 | -0,70   | -1,56 | -0,96 | -0,90 | -1,29 | -0,48  | -0,96   | -1,19 |
| -1,48 | -0,45  | 0,57   | -1,06 | 0,22  | -1,05 | -1,00   | -0,64 | -1,05 | -0,90 | -0,03 | -1,00  | -0,17   | 0,11  |
| -1,48 | -1,37  | -1,66  | -0,44 | -1,72 | -0,70 | -0,32   | -0,52 | -0,79 | -0,05 | 0,47  | -1,59  | -2,12   | -1,79 |
| -0,49 | 0,10   | 0,57   | -0,35 | -0,25 | 0,27  | -0,53   | -0,08 | -0,64 | -0,90 | -0,54 | -0,35  | 0,09    | 0,11  |
| -0,98 | 0,08   | 0,20   | -0,06 | -1,54 | -0,46 | 1,36    | 0,07  | 0,26  | 1,71  | 0,47  | -0,42  | -0,80   | -0,48 |
| -0,98 | 1,32   | 1,50   | 0,38  | 0,43  | 0,71  | 0,58    | 1,37  | 1,08  | 1,60  | 1,81  | 0,79   | 0,73    | 0,83  |
| -0,49 | -0,08  | -0,36  | 0,25  | -0,68 | 0,80  | 0,07    | -0,28 | -1,14 | -0,37 | -0,45 | -0,52  | -0,75   | -0,48 |
| -0,74 | -0,32  | 0,66   | -0,97 | -0,08 | -0,08 | -0,24   | -0,99 | -0,67 | 0,22  | -0,20 | -1,00  | -0,12   | 0,83  |
| -0,74 | 0,40   | 1,22   | -0,51 | 0,05  | 0,13  | -1,07   | -0,37 | 0,17  | -0,37 | 0,30  | 0,86   | 0,99    | 0,11  |
| -1,23 | 0,75   | 1,03   | 0,08  | 1,77  | -0,08 | -0,23   | 0,56  | 0,23  | 1,44  | 1,31  | 0,76   | 0,41    | 1,30  |
| -0,49 | 0,17   | 1,41   | -0,94 | 0,22  | 0,22  | -0,92   | -0,75 | -0,76 | 0,38  | 1,48  | -0,21  | 0,09    | 0,47  |
| -0,24 | 1,60   | 1,96   | 0,31  | 0,61  | 0,56  | 0,09    | 1,35  | 1,13  | 0,11  | 2,06  | 1,31   | 0,52    | 1,07  |
| 1,00  | -0,10  | -0,45  | 0,30  | 0,18  | 0,30  | 0,19    | 0,51  | 0,32  | -0,37 | -1,63 | -0,31  | -0,33   | -0,48 |
| 0,51  | 1,08   | 0,20   | 1,29  | 0,61  | 0,64  | 1,61    | 1,55  | 1,08  | 1,34  | 0,22  | 1,21   | 1,25    | 0,83  |
| 1,75  | 0,25   | -0,45  | 0,79  | 1,08  | 0,19  | 0,86    | -0,33 | 1,95  | 0,00  | -0,29 | 0,93   | 0,67    | 0,11  |
| 1,75  | -0,95  | -1,76  | 0,30  | -0,68 | -0,16 | 0,36    | -0,52 | 0,23  | -0,16 | -1,46 | -0,80  | -1,49   | -1,43 |
| 1,50  | 0,63   | 0,10   | 0,77  | 0,39  | 0,29  | 0,86    | 0,80  | 0,73  | 0,96  | -0,29 | 0,96   | 0,83    | -0,96 |
| 1,00  | 1,27   | 0,10   | 1,64  | 1,77  | 1,21  | 1,01    | 1,20  | 1,34  | 0,54  | 0,81  | 1,66   | 1,25    | 1,42  |
| 0,01  | 0,05   | -1,11  | 1,16  | 0,00  | 1,08  | -0,09   | 0,18  | 0,55  | -0,16 | -0,45 | 1,31   | 0,67    | 0,47  |
| 1,25  | 1,39   | 1,03   | 0,89  | 0,86  | 0,98  | 1,43    | 1,64  | 1,02  | 0,86  | 0,47  | 1,28   | 0,46    | 0,11  |
| 0,01  | -0,17  | -1,01  | 0,74  | 0,48  | 0,51  | 0,79    | 0,73  | 0,35  | -0,53 | -1,12 | -0,31  | 0,57    | 0,35  |
| 0,01  | 0,86   | -0,36  | 1,55  | 1,42  | 1,82  | 1,19    | 0,71  | 0,32  | 0,86  | 0,97  | -0,14  | 1,89    | 2,14  |
| 0,76  | -2,03  | -0,92  | -2,06 | -1,76 | -2,28 | -1,92   | -1,54 | -1,43 | -1,97 | -1,04 | -1,90  | -1,75   | -1,67 |
| 1,00  | -1,62  | -0,27  | -1,98 | -1,07 | -2,17 | -1,42   | -1,80 | -2,16 | -1,49 | -1,12 | -1,24  | -1,33   | -1,19 |
| 0,01  | -1,50  | -1,11  | -1,14 | -1,50 | -1,60 | -1,94   | -1,27 | -1,14 | -1,86 | -0,45 | -0,80  | -0,64   | -0,60 |

**Gambar 4.18 Standarisasi Sampel Data**

Mengacu pada persamaan 2.1, penulis perlu mencari nilai *y-intercept* ( $\delta_0$ ), karena dilakukan standarisasi pada semua variabel independen (w) maka nilai *y-intercept* ( $\delta_0$ ) adalah nilai rata-rata dari variabel dependen ( $\bar{v}$ ) yaitu sebagai berikut:

$$\bar{v} = \frac{v_1 + v_2 + \dots + v_n}{n}$$

$$\bar{v} = \frac{12,3 + 6,1 + \dots + 14}{25}$$

$$\bar{v} = \frac{417}{25} = 16,68$$

Penulis sudah memiliki nilai  $\bar{v}$  sebesar 16,68 yang penulis gunakan sebagai  $\delta_0$ . Penulis juga sudah memiliki nilai koefisien ( $\delta_{1-n}$ ) yang dapat dilihat pada tabel 4.1 dan nilai variabel independen ( $w_{1-n}$ ) yang sudah distandarisasi dapat dilihat pada gambar 4.17, selanjutnya penulis memasukkan nilai tersebut pada persamaan 2.1.

|                       | Age   | Weight | Height | BMI   | Neck  | Chest | Abdomen | Hip   | Thigh | Knee  | Ankle | Biceps | Forearm | Wrist |
|-----------------------|-------|--------|--------|-------|-------|-------|---------|-------|-------|-------|-------|--------|---------|-------|
| Nilai Variabel        | -1,23 | -1,36  | -1,11  | -0,92 | -0,77 | -1,13 | -0,70   | -1,56 | -0,96 | -0,90 | -1,29 | -0,48  | -0,96   | -1,19 |
| Nilai Koefisien Ridge | 1,31  | 0,12   | -1,43  | -0,87 | -0,49 | 0,33  | 6,23    | -0,30 | 0,92  | 0,41  | 0,07  | 0,53   | 0,41    | -1,79 |
| Nilai Koefisien Lasso | 0,85  | 0,00   | -0,98  | -0,38 | -0,37 | -0,02 | 7,71    | -0,37 | 0,19  | 0,00  | 0,00  | 0,43   | 0,26    | -1,66 |

- Ridge Regression

$$v_{ridge} = 16,68 + (1,31x - 1,23) + (0,12x - 1,36) + \dots + (-1,79x - 1,19)$$

$$v_{ridge} = 16,68 + (-1,61) + (-0,16) + \dots + (2,14)$$

$$v_{ridge} = 16,68 + (-3,14)$$

$$v_{ridge} = \mathbf{13,54}$$

- LASSO Regression

$$v_{lasso} = 16,68 + (0,85x - 1,23) + (0,00x - 1,36) + \dots + (-1,66x - 1,19)$$

$$v_{lasso} = 16,68 + (-1,61) + (-0,16) + \dots + (2,14)$$

$$v_{lasso} = 16,68 + (-2,77)$$

$$v_{lasso} = \mathbf{13,91}$$

Selanjutnya penulis menambahkan nilai penalti koefisien menggunakan persamaan 2.3 untuk *Ridge Regression* dan persamaan 2.6 untuk *LASSO Regression*.

- Ridge Regression

$$\lambda_{ridge} \sum ((1,31^2) + (0,12^2) + \dots + (-1,79^2))$$

$$\lambda_{ridge} \sum (1,72) + (0,01) + \dots + (3,20))$$

$$\lambda_{ridge} (48,46)$$

- LASSO Regression

$$\lambda_{lasso} \sum |0,85| + |0,00| + \dots + |-1,66|$$

$$\lambda_{lasso} \sum 0,85 + 0,00 + \dots + 1,66$$

$$\lambda_{lasso} (13,22)$$

Selanjutnya masukkan nilai yang sudah didapat dan nilai lambda sebesar 0,1 ke dalam persamaan 2.4 untuk *Ridge Regression* dan persamaan 2.7 untuk *LASSO Regression*.

- Ridge Regression

$$v_{pred\ ridge} = 16,68 + (-3,14) + (0,1 \times 48,46)$$

$$v_{pred\ ridge} = 13,54 + (4,84)$$

$$v_{pred\ ridge} = \mathbf{18,38}$$

- LASSO Regression

$$v_{pred\ lasso} = 16,68 + (-2,77) + (0,1 \times 13,22)$$

$$v_{pred\ lasso} = 13,91 + (1,32)$$

$$v_{pred\ lasso} = \mathbf{15,23}$$

Selanjutnya penulis menghitung nilai error dengan mengurangi nilai *bodyfat* aktual dikurangi nilai *bodyfat* prediksi. Sebagai berikut:

| Aktual | Prediksi Ridge | Prediksi LASSO | Error Ridge | Error LASSO |
|--------|----------------|----------------|-------------|-------------|
| 12,3   | 18,38          | 15,23          | -6,08       | -2,93       |



#### 4.4. Pengujian Model

Hasil dari prediksi persentase lemak tubuh untuk kedua algoritma yaitu *Ridge* dan *LASSO Regression* dapat dilihat pada tabel 4.2 ini.

**Tabel 4.2** Hasil Prediksi

| Actual | Predicted (Ridge) | Predicted (Lasso) | Error (Ridge) | Error (Lasso) |
|--------|-------------------|-------------------|---------------|---------------|
| 23,3   | 19,69328          | 19,61582          | 3,606719      | 3,684184      |
| 19,2   | 17,46119          | 17,72273          | 1,738807      | 1,477274      |
| 11,3   | 17,72784          | 16,03192          | -6,42784      | -4,73192      |
| 26,1   | 25,39885          | 25,6619           | 0,701149      | 0,438103      |
| 22,1   | 19,1728           | 19,14588          | 2,927196      | 2,95412       |
| 11,5   | 15,51468          | 15,38527          | -4,01468      | -3,88527      |
| 6,6    | 8,526473          | 7,011117          | -1,92647      | -0,41112      |
| 8,6    | 12,47231          | 12,07249          | -3,87231      | -3,47249      |
| 11,7   | 12,73711          | 12,1855           | -1,03711      | -0,4855       |
| 25,4   | 22,86669          | 24,23371          | 2,533311      | 1,166287      |
| 22,1   | 15,17259          | 15,64893          | 6,927414      | 6,451073      |
| 24,5   | 23,79172          | 23,64233          | 0,708282      | 0,857665      |
| 12,4   | 15,60365          | 13,38013          | -3,20365      | -0,98013      |
| 30,4   | 30,44675          | 30,07379          | -0,04675      | 0,326214      |
| 13,8   | 14,20372          | 14,79074          | -0,40372      | -0,99074      |
| 14,8   | 19,17528          | 18,77055          | -4,37528      | -3,97055      |
| 25,5   | 21,60773          | 21,81011          | 3,892268      | 3,68989       |
| 20,9   | 22,78538          | 22,28295          | -1,88538      | -1,38295      |
| 29,9   | 30,85322          | 30,73811          | -0,95322      | -0,83811      |
| 14     | 9,229842          | 8,197925          | 4,770158      | 5,802075      |
| 29,3   | 34,99476          | 36,29286          | -5,69476      | -6,99286      |
| 16,5   | 23,99441          | 24,03712          | -7,49441      | -7,53712      |
| 26,7   | 20,11812          | 20,27364          | 6,581876      | 6,426358      |
| 6,3    | 11,16684          | 9,890493          | -4,86684      | -3,59049      |
| 13,1   | 16,79459          | 15,61804          | -3,69459      | -2,51804      |
| 13,9   | 10,83613          | 9,966922          | 3,063867      | 3,933078      |

|      |          |          |          |          |
|------|----------|----------|----------|----------|
| 22,6 | 23,17423 | 24,21862 | -0,57423 | -1,61862 |
| 7,1  | 10,59135 | 9,932669 | -3,49135 | -2,83267 |
| 11,9 | 14,08603 | 14,12352 | -2,18603 | -2,22352 |
| 27,9 | 20,37787 | 21,28343 | 7,522127 | 6,616568 |
| 28,7 | 25,96777 | 26,04074 | 2,732225 | 2,659257 |
| 21,3 | 26,15711 | 24,53823 | -4,85711 | -3,23823 |
| 18,3 | 20,50487 | 20,56222 | -2,20487 | -2,26222 |
| 26   | 30,81551 | 30,75585 | -4,81551 | -4,75585 |
| 3,7  | 8,84579  | 9,114547 | -5,14579 | -5,41455 |
| 28   | 29,95449 | 30,38983 | -1,95449 | -2,38983 |
| 16   | 18,3128  | 16,99495 | -2,3128  | -0,99495 |
| 17,7 | 19,41377 | 19,10806 | -1,71377 | -1,40806 |
| 26,6 | 25,69385 | 25,58537 | 0,906151 | 1,014633 |
| 22,5 | 17,59503 | 17,99223 | 4,90497  | 4,50777  |
| 25,8 | 18,3852  | 17,70034 | 7,414803 | 8,099661 |
| 9,6  | 16,96624 | 16,98427 | -7,36624 | -7,38427 |
| 17,3 | 14,23096 | 13,58683 | 3,069045 | 3,713167 |
| 23,6 | 27,32175 | 27,73211 | -3,72175 | -4,13211 |
| 10,9 | 20,42755 | 20,73961 | -9,52755 | -9,83961 |
| 29   | 24,2052  | 23,568   | 4,794805 | 5,431997 |
| 19,2 | 22,88067 | 21,83619 | -3,68067 | -2,63619 |
| 12,9 | 14,36623 | 14,76251 | -1,46623 | -1,86251 |
| 16,9 | 16,94886 | 17,38126 | -0,04886 | -0,48126 |
| 35,2 | 51,88974 | 52,76886 | -16,6897 | -17,5689 |

Model yang telah dilatih menggunakan data latih dan di uji menggunakan data uji dilakukan pengujian model dengan matrik evaluasi, seperti MSE, MAE,  $R^2$ , dan Adjusted  $R^2$ . Berikut merupakan perbandingan nilai matrik evaluasi untuk tiap-tiap algoritma pada data latih dan data uji.

**Tabel 4.3** Matriks Evaluasi

| Matriks Evaluasi        | Ridge    |         | LASSO    |         |
|-------------------------|----------|---------|----------|---------|
|                         | Training | Testing | Training | Testing |
| MSE                     | 17.876   | 22.895  | 17.666   | 22.567  |
| MAE                     | 3.496    | 3.808   | 3.475    | 3.641   |
| R <sup>2</sup>          | 0.736    | 0.597   | 0.739    | 0.603   |
| Adjusted R <sup>2</sup> | 0.716    | 0.437   | 0.719    | 0.445   |

a. MSE (*Mean Squared Error*)

MSE atau kesalahan kuadrat rata-rata yang juga disebut deviasi kuadrat rata-rata adalah kuadrat perbedaan antara nilai aktual dan prediksi nilai. MSE memberi tahu penulis seberapa dekat garis kecocokan terbaik ke sekumpulan titik. MSE selalu bernilai positif. Kuadrat-kuadrat diambil untuk menghilangkan tanda-tanda negatif. Semakin dekat MSE mendekati 0, semakin akurat prediksinya. Berikut rumus MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_{aktual} - \hat{Y}_{prediksi})^2 \quad (4.1)$$

Berdasarkan tabel 4.1 yang memuat hasil prediksi maka MSE dapat dinyatakan sebagai berikut:

- Ridge Regression

$$\begin{aligned}
 &= (Y_{akt,1} - Y_{pred,1})^2 + (Y_{akt,2} - Y_{pred,2})^2 + \dots + (Y_{akt,n} - Y_{pred,n})^2 \\
 &= (23,30 - 19,69)^2 + (19,20 - 17,46)^2 + \dots + (35,20 - 51,88)^2 \\
 &= (3,61)^2 + (1,74)^2 + \dots + (-16,68)^2 \\
 &= (13,03) + (3,02) + \dots + (278,22) \\
 &= 1144,77
 \end{aligned}$$

$$MSE_{ridge} = \frac{1144,77}{50} = \mathbf{22,89}$$

- LASSO Regression

$$\begin{aligned}
 &= (Y_{akt,1} - Y_{pred,1})^2 + (Y_{akt,2} - Y_{pred,2})^2 + \dots + (Y_{akt,n} - Y_{pred,n})^2 \\
 &= (23,30 - 19,61)^2 + (19,20 - 17,72)^2 + \dots + (35,20 - 52,76)^2
 \end{aligned}$$

$$\begin{aligned}
&= (3,69)^2 + (1,48)^2 + \dots + (-17,56)^2 \\
&= (13,61) + (2,19) + \dots + (308,35) \\
&= 1128,40
\end{aligned}$$

$$MSE_{lasso} = \frac{1128,40}{50} = \mathbf{22,56}$$

b. MAE (*Mean Absolute Error*)

MAE atau kesalahan absolut rata-rata yang juga disebut deviasi mean absolute memberikan penulis rata-rata perbedaan absolut antara antara nilai yang diamati dan nilai yang diprediksi. Perbedaannya antara MAE dan MSE adalah bahwa MAE mengambil nilai absolut perbedaan antara nilai yang diprediksi dan yang diamati sedangkan MSE mengambil perbedaan kuadrat. Berikut rumus MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_{aktual} - \hat{Y}_{prediksi}| \quad (4.2)$$

Berdasarkan tabel 4.1 yang memuat hasil prediksi maka MAE dapat dinyatakan sebagai berikut:

- Ridge Regression

$$\begin{aligned}
&= |Y_{akt,1} - Y_{pred,1}| + |Y_{akt,2} - Y_{pred,2}| + \dots + |Y_{akt,n} - Y_{pred,n}| \\
&= |23,30 - 19,69| + |19,20 - 17,46| + \dots + |35,20 - 51,88| \\
&= |3,61| + |1,74| + \dots + |-16,68| \\
&= 3,61 + 1,74 + \dots + 16,68 \\
&= 190,44
\end{aligned}$$

$$MAE_{ridge} = \frac{190,44}{50} = \mathbf{3,80}$$

- LASSO Regression

$$\begin{aligned}
&= |Y_{akt,1} - Y_{pred,1}| + |Y_{akt,2} - Y_{pred,2}| + \dots + |Y_{akt,n} - Y_{pred,n}| \\
&= |23,30 - 19,61| + |19,20 - 17,72| + \dots + |35,20 - 52,76| \\
&= |3,69| + |1,48| + \dots + |-17,56| \\
&= 3,69 + 1,48 + \dots + 17,56 \\
&= 182,08
\end{aligned}$$



$$MAE_{lasso} = \frac{182,08}{50} = 3,64$$

c.  $R^2$  (*R-Squared*)

$R^2$  disebut juga koefisien determinasi.  $R^2$  menentukan proporsi varians dalam variabel dependen yang dapat dijelaskan oleh variabel independen. R-Squared memberi penulis kecocokan (sejauh mana nilai yang diamati sesuai dengan nilai prediksi) untuk model regresi. Berikut rumus  $R^2$ :

$$R^2 = 1 - \frac{SSR \text{ (Sum of Squares Residual)}}{TSS \text{ (Total Sum of Squares)}} \quad (4.3)$$

$$SSR = \sum_{i=1}^n (Y_{aktual} - Y_{prediksi})^2 \quad (4.4)$$

$$TSS = \sum_{i=1}^n (Y_{aktual} - \bar{Y}_{aktual})^2 \quad (4.5)$$

Berdasarkan tabel 4.1 yang memuat hasil prediksi maka  $R^2$  dapat dinyatakan sebagai berikut:

- Menghitung TTS (*Total Sum of Squares*)

$$\begin{aligned} &= (Y_{akt,1} - \bar{Y}_{akt})^2 + (Y_{akt,2} - \bar{Y}_{akt})^2 + \dots + (Y_{akt,n} - \bar{Y}_{akt})^2 \\ &= (23,30 - 19,17)^2 + (19,20 - 19,17)^2 + \dots + (35,20 - 19,17)^2 \\ &= (4,13)^2 + (0,03)^2 + \dots + (16,03)^2 \\ &= 17,05 + 0,00 + \dots + 256,96 \\ TTS &= \mathbf{2847,46} \end{aligned}$$

- Menghitung SSR (*Sum of Squares Residual*)

$$\begin{aligned} SSR &= (Y_{akt,1} - Y_{pred,1})^2 + (Y_{akt,2} - Y_{pred,2})^2 + \dots + (Y_{akt,n} - Y_{pred,n})^2 \\ SSR_{ridge} &= \mathbf{1144,77} \\ SSR_{lasso} &= \mathbf{1128,40} \end{aligned}$$

- Ridge Regression

$$R^2_{ridge} = 1 - \frac{1144,77}{2847,46}$$



$$R^2_{ridge} = 1 - 0,402 = \mathbf{0,59}$$

- LASSO Regression

$$R^2_{lasso} = 1 - \frac{1128,40}{2847,46}$$

$$R^2_{lasso} = 1 - 0,396 = \mathbf{0,60}$$

d. *Adjusted R<sup>2</sup>*

*Adjusted R<sup>2</sup>* juga memberikan kesesuaian seperti *R<sup>2</sup>*. Perbedaannya adalah bahwa *Adjusted R<sup>2</sup>* disesuaikan dengan jumlah prediktor untuk model. Nilai dari *Adjusted R<sup>2</sup>* akan menurun seiring dengan bertambahnya jumlah variabel yang tidak relevan.

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \quad (4.6)$$

Dimana:

n = jumlah data

p = jumlah variabel independen

- Ridge Regression

$$R^2_{adj ridge} = 1 - \frac{(1 - 0,597)(50 - 1)}{50 - 14 - 1}$$

$$R^2_{adj ridge} = 1 - \frac{(0,403)(49)}{35}$$

$$R^2_{adj ridge} = 1 - \frac{19,74}{35}$$

$$R^2_{adj ridge} = 1 - 0,564 = \mathbf{0,43}$$

- LASSO Regression

$$R^2_{adj lasso} = 1 - \frac{(1 - 0,603)(50 - 1)}{50 - 14 - 1}$$

$$R^2_{adj lasso} = 1 - \frac{(0,397)(49)}{35}$$

$$R^2_{adj lasso} = 1 - \frac{19,45}{35}$$

$$R^2_{adj lasso} = 1 - 0,555 = \mathbf{0,44}$$

#### 4.5. Hasil Pengujian

Ditarik kesimpulan, bahwa model *LASSO Regression* lebih unggul untuk keseluruhan matrik evaluasi, dengan nilai MSE sebesar 22.56, MAE sebesar 3.64, R-squared sebesar 0.6, dan Adjusted R-squared sebesar 0.44, berbanding terbalik dengan Ridge Regression dengan nilai MSE hanya sebesar 22.89, MAE sebesar 3.80, R-squared sebesar 0.59, dan Adjusted R-squared sebesar 0.43, sehingga penulis menggunakan model *LASSO Regression* pada implementasi web.

**Tabel 4.4** Perbandingan Nilai Matrik Evaluasi

| Matrik Evaluasi | Ridge | LASSO |
|-----------------|-------|-------|
| MSE             | 22.89 | 22.56 |
| MAE             | 3.80  | 3.64  |
| $R^2$           | 0.59  | 0.60  |
| Adjusted $R^2$  | 0.43  | 0.44  |

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1. Kesimpulan**

Berdasarkan hasil penelitian maupun pengujian yang sudah dilakukan oleh penulis terhadap prediksi persentase lemak tubuh menggunakan algoritma *Ridge* dan *LASSO Regression*, maka didapat beberapa kesimpulan:

1. Hasil penelitian menunjukkan bahwa *LASSO Regression* sedikit lebih unggul dibanding *Ridge Regression*, dengan MSE sebesar 22.56, MAE sebesar 3.64, R-squared sebesar 0.60, dan Adjusted R-squared sebesar 0.44. Data yang digunakan berjumlah 250 baris, yang dibagi menjadi dua bagian berdasarkan persentase 80:20 dari total data. Sebanyak 200 baris data (80%) digunakan sebagai data latih, sedangkan 50 baris data (20%) digunakan sebagai data uji.
2. Penerapan *Ridge* dan *LASSO Regression* dalam bidang kesehatan, khususnya prediksi persentase lemak tubuh, dapat membantu praktisi kesehatan dan peneliti dalam mengetahui persentase lemak tubuh dengan cepat dan akurat. Dengan demikian, tindakan pencegahan dan penanganan dapat dilakukan lebih awal untuk mengurangi risiko penyakit yang dapat ditimbulkan. Penelitian ini memberikan kontribusi positif dalam upaya pencegahan dan penanganan obesitas dan lemak tubuh berlebih.

#### **5.2. Saran**

Berdasarkan temuan dari studi yang telah dilakukan, peneliti mengusulkan bahwa penelitian berikutnya dapat:

1. Menggunakan data yang lebih variatif, seperti menggunakan data asli atau menambahkan data untuk wanita.
2. Melakukan eliminasi variabel independen, mengingat banyaknya variabel yang digunakan tetapi hanya sedikit yang berpengaruh besar terhadap variabel dependen.

## DAFTAR PUSTAKA

- Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*, 1142(1). <https://doi.org/10.1088/1742-6596/1142/1/012012>
- Enwere, K., Nduka, E., & Ogoke, U. (2023). Comparative Analysis of Ridge, Bridge and Lasso Regression Models In the Presence of Multicollinearity. *IPS Intelligentsia Multidisciplinary Journal*, 3(1), 1–8. <https://doi.org/10.54117/iimj.v3i1.5>
- Ferenci, T., & Kovács, L. (2018). Predicting body fat percentage from anthropometric and laboratory measurements using artificial neural networks. *Applied Soft Computing Journal*, 67, 834–839. <https://doi.org/10.1016/j.asoc.2017.05.063>
- Hasanah, U., & Nurhopipah, A. (2023). Prediksi Persentase Body Fat Menggunakan Algoritma CART dan M5'. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 4(4), 351–363. <https://doi.org/10.35746/jtim.v4i4.316>
- Huang, L., Teo, C. J. H., Luan, Y., & Tekwe, C. D. (2023). Predicting Body Fat Percentage from Anthropometric Measurements in Asian Athletes. *Journal of Science in Sport and Exercise*, 1–10.
- Jawad, M., & Mughal, H. (2018). Data Mining: Web Data Mining Techniques, Tools and Algorithms: An Overview. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 9, Issue 6). [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- Mahesh, B. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, 9(1), 381–386. <https://doi.org/10.21275/ART20203995>
- Ren, P. (2024). Comparison and analysis of the accuracy of Lasso regression, Ridge regression and Elastic Net regression models in predicting students' teaching

quality achievement. *Applied and Computational Engineering*, 51(1), 313–319.  
<https://doi.org/10.54254/2755-2721/51/20241625>

Sitanggang, D. (2022). Model Prediksi Obesitas Dengan Menggunakan Support Vector Machine. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima*), 5(2).

Tatachar, A. v. (2021). Comparative Assessment of Regression Models Based On Model Evaluation Metrics. *International Research Journal of Engineering and Technology*. [www.irjet.net](http://www.irjet.net)

Thevaraja, M., Rahman, A., & Gabirial, M. (2019). Recent developments in data science: Comparing linear, ridge and lasso regressions techniques using wine data. *Proceedings of the International Conference on Digital Image & Signal Processing*, 1–6.

World Health Organization. (2024, March 1). *Obesity and overweight*.  
<https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>

Zai, C., & Komputer, T. (2022). Implementasi Data Mining Sebagai Pengolahan Data. In *Portalddata.org* (Vol. 2, Issue 3).

