

TESIS

**ESTIMASI BOBOT SAPI MENGGUNAKAN LINEAR REGRESSION
DAN RANDOM FOREST REGRESSOR**



Disusun oleh:

Nama : Anjar Setiawan
NIM : 22.51.2266
Konsentrasi : Business Intelligence

**PROGRAM STUDI S2 INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**



2024

TESIS

**ESTIMASI BOBOT SAPI MENGGUNAKAN LINEAR REGRESSION
DAN RANDOM FOREST REGRESSOR**

**CATTLE WEIGHT ESTIMATION USING LINEAR REGRESSION
AND RANDOM FOREST REGRESSOR**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Anjar Setiawan
NIM : 22.51.2266
Konsentrasi : Business Intelligence

**PROGRAM STUDI S2 INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

HALAMAN PENGESAHAN

**ESTIMASI BOBOT SAPI MENGGUNAKAN LINEAR REGRESSION DAN
RANDOM FOREST REGRESSOR**

**CATTLE WEIGHT ESTIMATION USING LINEAR REGRESSION AND
RANDOM FOREST REGRESSOR**

Dipersiapkan dan Disusun oleh

Anjar Setiawan

22.51.2266

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 08 Mei 2024

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer.

Yogyakarta, 08 Mei 2024

Rektor

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

ESTIMASI BOBOT SAPI MENGGUNAKAN LINEAR REGRESSION DAN RANDOM FOREST REGRESSOR

CATTLE WEIGHT ESTIMATION USING LINEAR REGRESSION AND RANDOM FOREST REGRESSOR

Dipersiapkan dan Disusun oleh

Anjar Setiawan

22.51.2266

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Informatika
Program Pascasarjana Universitas AMIKOM
Yogyakarta pada hari Rabu, 08 Mei 2024

Pembimbing Utama

Prof. Dr. Ema Utami, S.Si, M.Kom
NIK. 190302037

Anggota Tim Penguji

Dr. Andi Sunvoto, M.Kom
NIK. 190302052

Pembimbing Pendamping

Alva Hendi Muhammad, S.T., M.Eng., Ph.D
NIK. 190302493

Dhani Ariatmanto, M.Kom., Ph.D.
NIK. 190302197

Prof. Dr. Ema Utami, S.Si, M.Kom
NIK. 190302037

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 08 Mei 2024
Direktur Program Pascasarjana

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Anjar Setiawan
NIM : 22.51.2266
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:

Estimasi Bobot Sapi Menggunakan Linear Regression dan Random Forest Regressor

Dosen Pembimbing Utama : Prof. Dr. Ema Utami, S.Si., M.Kom.
Dosen Pembimbing Pendamping : Dhani Ariatmanto, M.Kom., Ph.D.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 08 Mei 2024

Yang Menyatakan,



Anjar Setiawan

HALAMAN PERSEMBAHAN

Kupersembahkan kepada ibu dan bapak yang telah mendidik penulis dengan penuh kesabaran dan kasih sayang. Keluarga besar yang selalu mendukung proses perjuangan penulis dengan penuh kasih sayang dan semangat daya juang yang tinggi. Harus berani, percaya diri, pantang menyerah, daya juang yang tinggi, motivasi tinggi, pandai bergaul, komunikatif, adaptif, akhlak dan sikap yang baik. Kepada ibu dan bapak rindho dan keberkahanmu adalah sumber kekuatan dan kesuksesan bagi penulis.



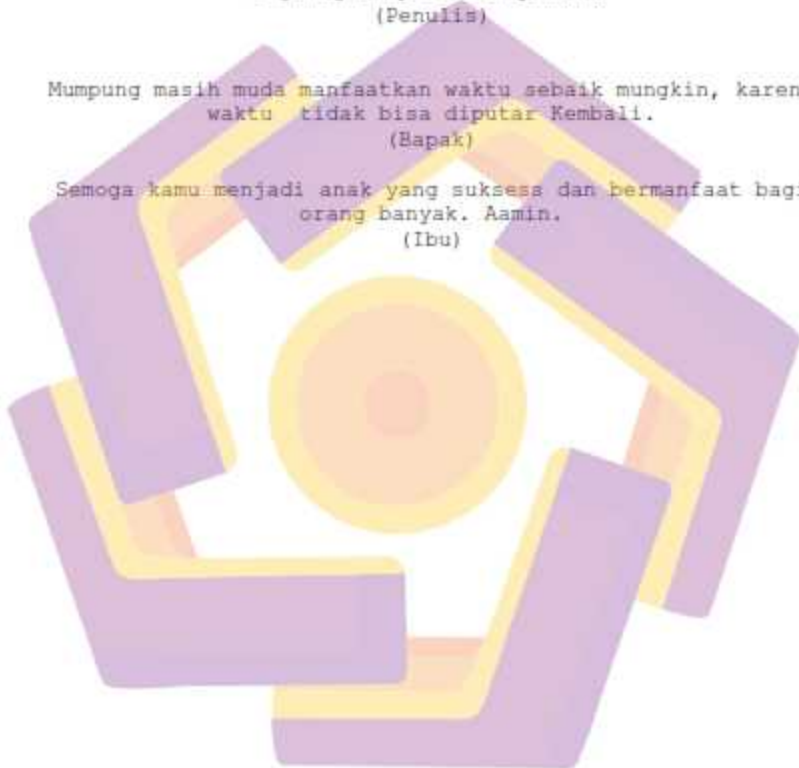
HALAMAN MOTTO

Never give up. Today is hard, tomorrow will be even harder,
but the day after tomorrow will be beautiful.
(Jack Ma)

Pejuang mimpi tersenyumlah.
(Penulis)

Mumpung masih muda manfaatkan waktu sebaik mungkin, karena
waktu tidak bisa diputar kembali.
(Bapak)

Semoga kamu menjadi anak yang sukses dan bermanfaat bagi
orang banyak. Aamin.
(Ibu)



KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT yang telah memberikan rahmat-Nya sehingga penulis dapat menyelesaikan laporan penelitian tesis ini dengan baik. Penulisan laporan tesis ini dapat terselesaikan berkat bantuan dari berbagai pihak. Oleh karena itu penulis mengucapkan terimakasih kepada pihak-pihak yang terlibat dalam penelitian ini:

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom. selaku Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta.
3. Ibu Prof. Dr. Ema Utami, S.Si., M.Kom. selaku Wakil Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta sekaligus selaku Pembimbing Utama.
4. Bapak Dhani Ariatmanto, M.Kom., Ph.D. selaku dosen Pembimbing Pendamping.
5. Bapak Dr. Andi Sunyoto, M.Kom. selaku penguji Seminar Ujian Tesis.
6. Bapak Alva Hendi Muhammad, S.T., M.Eng., Ph.D. selaku penguji Seminar Ujian Tesis.
7. Ibu Prof. Dr. Ema Utami, S.Si., M.Kom. selaku penguji Seminar Ujian Tesis.
8. Serta semua pihak yang tidak bisa penulis sebutkan satu persatu terima kasih atas doa dan sarannya.

Dengan diiringi doa dan ucapan terimakasih, penulis berharap semoga tesis ini dapat bermanfaat. Saran, harapan, dan kritik yang membangun. Terimakasih

Yogyakarta, 08 Mei 2024

Penulis

DAFTAR ISI

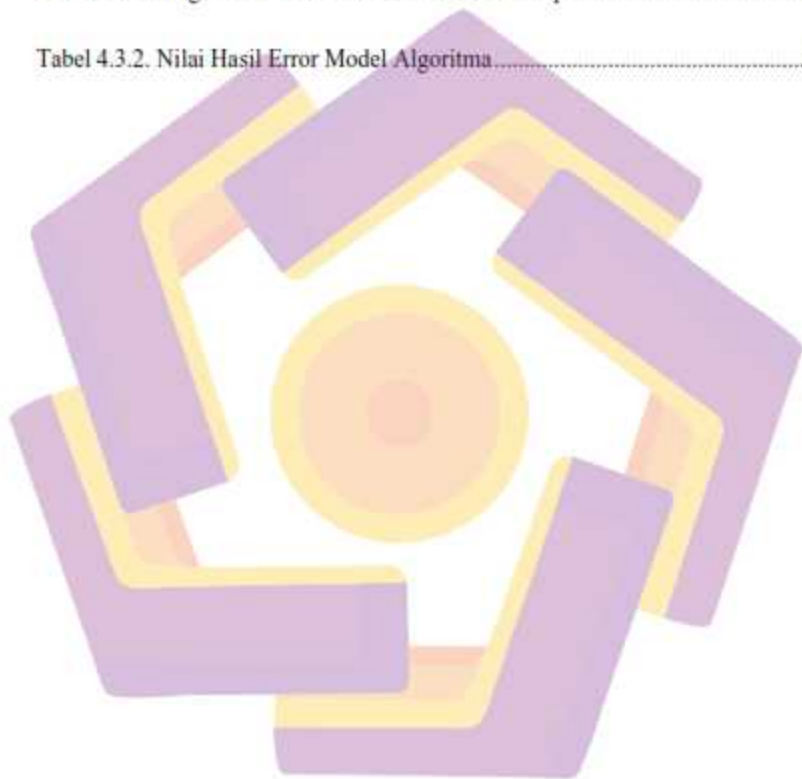
HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PERNYATAAN KEASLIAN TESIS	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR ISTILAH	xv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	6
1.3. Batasan Masalah	6
1.4. Tujuan Penelitian	7
1.5. Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA	8
2.1. Tinjauan Pustaka	8

2.2. Keaslian Penelitian.....	11
2.3. Landasan Teori.....	14
2.3.1. Bobot Sapi.....	14
2.3.2. Machine Learning.....	15
2.3.3. Random Forest Regressor.....	18
2.3.4. Linear Regression.....	20
2.3.5. Mean Absolute Error (MAE).....	22
2.3.6. Root Mean Square Error (RMSE).....	23
2.3.7. R square (R2).....	25
2.3.8. Mean Absolute Percentage Error (MAPE).....	26
BAB III METODE PENELITIAN.....	28
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	28
3.2. Metode Pengumpulan Data.....	28
3.3. Metode Analisis Data.....	30
3.4. Dataset.....	30
3.5. Alur Penelitian.....	35
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	38
4.1. Dataset Penelitian.....	38
4.1.1. Preprocessing Data.....	38
4.2. Pemodelan Linear Regression dan Random Forest Regressor.....	43
4.3. Hasil dan Pembahasan.....	46
4.3.1. Correlation Heatmap.....	46

4.3.2. Data Normalization.....	47
4.3.3. Feature Selection.....	49
4.3.4. K-Fold Cross Validation.....	50
4.3.5. Algoritma Linear Regression.....	51
4.3.6. Algoritma Random Forest Regressor.....	55
4.3.7. Nilai Hasil Model Algoritma.....	58
BAB V PENUTUP.....	61
5.1. Kesimpulan.....	61
5.2. Saran.....	62
DAFTAR PUSTAKA.....	64
LAMPIRAN 1. DATASET SAPI.....	71

DAFTAR TABEL

Tabel 2. 1. Matriks Literatur Review dan Posisi Penelitian	11
Tabel 3. 1. Cuplikan Dataset Sapi Hereford	34
Tabel 4.3.1. Singkatan dan Definisi Ukuran Tubuh Sapi	41
Tabel 4.3.2. Nilai Hasil Error Model Algoritma	58



DAFTAR GAMBAR

Gambar 2. 1. Hewan Ternak Sapi.....	15
Gambar 2. 2. Alur Kerja Machine Learning.....	17
Gambar 2. 4. Cara Kerja Random Forest Regressor.....	19
Gambar 2. 5. Garis Linear Regression.....	21
Gambar 3. 1. Sapi Hereford.....	29
Gambar 3. 2. Cuplikan isi dataset sapi Hereford.....	32
Gambar 3. 3. Alur Penelitian.....	35
Gambar 4. 1. Dimensi Tubuh Sapi.....	38
Gambar 4. 2. <i>Correlation Heatmap</i>	46
Gambar 4. 3. <i>Pseudocode Data Normalization</i>	48
Gambar 4. 4. Cuplikan <i>Data Normalization</i>	48
Gambar 4. 5. <i>Pseudocode Feature Selection</i>	49
Gambar 4. 6. <i>Pseudocode K-Fold Cross-Validation</i>	50
Gambar 4. 7. Kinerja Metrik Algoritma Linear Regression dengan Sembilan Variabel dan 10-Fold Cross Validation.....	52
Gambar 4. 8. <i>Pseudocode Model Algorithm Linear Regression</i>	54
Gambar 4. 9. Kinerja Metrik Algoritma Random Forest Regressor dengan Sembilan Variabel dan 10-Fold Cross Validation.....	56
Gambar 4. 10. <i>Pseudocode Model Algorithm Random Forest Regressor</i>	57

DAFTAR ISTILAH

Model: Istilah umum yang digunakan untuk merepresentasikan kecerdasan buatan yang sudah dilatih.

Data Labelling: Sebuah proses untuk memberikan nama disetiap variabel pengukuran sapi. Tujuannya adalah untuk pengelompokan, pemodelan, dan analisis statistik yang lebih fokus dan relevan.

Data Cleaning: Sebuah proses untuk menjamin kualitas data. Tujuannya adalah untuk menghilangkan data yang tidak valid, tidak lengkap, dan tidak relevan. Ini juga memberikan hasil penelitian yang akurat dan dapat diandalkan.

Data Normalization: Sebuah proses untuk mengubah data menjadi bentuk standar, sehingga memudahkan dalam mengolah dan menganalisis data ternak. Normalisasi data bertujuan untuk memastikan bahwa setiap atribut mempunyai kontribusi yang seimbang untuk memperoleh hasil penelitian yang lebih akurat.

Fitur Selection: Sebuah proses untuk mengidentifikasi subset fitur yang paling relevan dan signifikan dalam dataset ternak. Pemilihan fitur bertujuan untuk mereduksi dimensi data, meningkatkan efisiensi komputasi, menghilangkan komponen redundant, dan meningkatkan performa model estimasi.

K-fold Cross Validation: Proses untuk menguji performa model secara lebih akurat dan andal dengan membagi data menjadi k subset dengan ukuran yang sama. Tujuan dari validasi silang k-fold juga untuk membantu mengevaluasi stabilitas dan generalisasi model pada data sapi yang belum pernah dilihat sebelumnya.

INTISARI

Industri peternakan sapi global memiliki manfaat sebagai sumber pangan, mata pencaharian, kontribusi ekonomi, pemulihan lingkungan lahan, dan sumber energi. Pentingnya estimasi bobot sapi bagi peternak adalah untuk memantau perkembangan hewan. Sedangkan bagi pedagang, mengetahui berat hewan memudahkan dalam menghitung harga daging hewan yang dibelinya. Dalam penelitian ini, penulis mengusulkan untuk memperkirakan bobot sapi dengan menggunakan regresi linier dan regresi hutan acak.

Dataset yang digunakan dalam penelitian ini terdiri dari tiga belas variabel yaitu berat badan hidup, umur sapi, tinggi badan layu, tinggi sakrum, kedalaman dada, lebar dada, lebar maclock, lebar sendi panggul, panjang badan miring, panjang punggung miring, lingkaran dada, ketebalan metakarpus, dan setengah ketebalan bagian belakang. Untuk mengetahui model yang menghasilkan nilai error MAE, MAPE, RMSE, dan R-Squared terkecil.

Selama fase evaluasi, penulis menggunakan metrik evaluasi data pelatihan seperti MAE, MAPE, RMSE, dan R-Squared untuk mengukur performa model. Hasil penelitian menunjukkan bahwa algoritma random forest regressor mampu menghasilkan nilai estimasi bobot sapi dengan performa terbaik. Model ini menghasilkan nilai mean absolute error (MAE) sebesar 21.902 kg, mean absolute perception error (MAPE) sebesar 4.201%, root mean squared error (RMSE) sebesar 29.433 kg, dan R² sebesar 0.761.

Kata kunci: sapi, machine learning, regresi linier, random forest regressor, model estimasi

ABSTRACT

The global cattle farming industry has benefits as a food source, livelihood, economic contribution, land environmental restoration, and energy source. The importance of estimating cow weight for farmers is to monitor animal development. Meanwhile, for traders, knowing the animal's weight makes it easier to calculate the price of the animal meat they buy. In this study, the authors propose to estimate the weight of cattle using linear regression and random forest regression.

The dataset used in this research consists of thirteen variables, namely live body weight, cow age, withers height, sacrum height, chest depth, chest width, clock width, hip joint width, oblique body length, oblique back length, chest circumference, the thickness of the metacarpus, and half the thickness of the back. To find the model that produces the smallest MAE, MAPE, RMSE, and R-squared error values.

The authors used training data evaluation metrics such as MAE, MAPE, RMSE, and R-squared during the evaluation phase to measure model performance. The research results show that the random forest regressor algorithm can produce estimated weight values for cattle with the best performance. This model produces a mean absolute error (MAE) of 21,902 kg, a mean absolute perception error (MAPE) of 4.201%, a root mean squared error (RMSE) of 29,433 kg, and an R2 of 0.761.

Keyword: cattle, machine learning, linear regression, random forest regressor, estimation model

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Industri peternakan sapi di Indonesia masih perlu ditingkatkan, hal ini terkait dengan kebutuhan produksi daging sapi. Saat ini 78% produksi daging sapi diperoleh dari dalam negeri, sisanya 5% diimport berupa daging sapi dan 17% sapi hidup (Soehadji, dalam Saleh et al. 2014). Memiliki kebermanfaatan sebagai sumber pangan, pekerjaan penghidupan, kontribusi ekonomi, pemulihan lahan lingkungan, dan sumber energi (Food and Agriculture., 2011). Hewan ternak sapi menjadi sumber daya penting, berkontribusi terhadap praktik pertanian, mendukung konservasi keanekaragaman hayati, memfasilitasi penelitian, dan memiliki makna budaya (Weber, C. et al., 2020).

Parameter kesehatan sapi seperti berat badan, suhu tubuh, denyut nadi, pernafasan, kondisi tubuh, dan produksi susu dapat digunakan untuk memantau pertumbuhan dan kesehatan hewan (Marai at al., 2007). Kualitas daging sapi sangat berkaitan dengan usia dan berat badan sapi (Roel Dohmen et al., 2021). Berdasarkan nilai pasar tertinggi saat produksi daging sapi pada usia prima yaitu antara 18 bulan hingga 24 bulan karena telah mencapai bobot optimal dan telah mengembangkan massa otot sapi (S. Paengkoum et al., 2021). Dalam konteks bisnis jual beli peternak adalah untuk membantu peternak dalam mengambil keputusan yang lebih baik terkait penjualan, pembelian hewan ternak, manajemen pakan, perawatan kesehatan, dan pemeliharaan ternak secara efisien (Weber, C. et

al. 2020). Kebutuhan hewan kurban pada Iduladha 2023 diproyeksikan mencapai 1,7 juta ekor. Jumlah sapi menempati urutan ke dua setelah kambing yaitu sebesar 650.282 ekor, kambing 743.672 ekor, domba 332.770 ekor, dan kerbau 16.327 ekor (Kementan, 2023).

(Institute for Demographic and Poverty Studies, 2023), juga menaksir kebutuhan hewan kurban pada Iduladha 2023 sebanyak 1,78 juta ekor. Jumlah itu terdiri dari 505.000 sapi dan kerbau serta 1,23 juta kambing dan domba. Dari jumlah tersebut, daging kurban yang bisa didapatkan sebanyak 103.000 ton.

Pentingnya menentukan daging sapi yang layak dikonsumsi adalah kualitas daging, kebersihan, kesehatan hewan, penyimpanan, pengangkutan, dan labelisasi dalam memenuhi standar keamanan dan kualitas pangan (Food Safety and Inspection Service, 2021). Mengetahui bertambahnya bobot sapi dilakukan dengan cara mengukur tinggi badan, panjang badan dan lingkar dada (J. Patel Ashwiniv et al., 2019). Estimasi bobot hewan menjadi sangat penting untuk peternak dalam rangka memonitor perkembangan hewan. Sedangkan untuk pedagang mengetahui bobot hewan memudahkan perhitungan harga dari daging hewan yang dibeli. Beberapa penelitian telah menerapkan penggunaan *machine learning* (ML) dan *deep learning* (DL) untuk mengestimasi bobot hewan sebagai cara inovasi teknologi.

Hasil penelitian (A N Ruchay., et al 2022), dengan menggunakan algoritma pembelajaran mesin terbaru, khususnya ExtraTreesRegressor, penelitian ini berhasil mengungkap bahwa metode ini mampu memberikan estimasi berat badan sapi Hereford yang paling akurat. Hasilnya, diukur dengan

menggunakan beberapa metrik evaluasi kinerja, menunjukkan tingkat akurasi yang luar biasa, dengan Mean Absolute Error (MAE) sebesar 23.154, Root Mean Squared Error (RMSE) sebesar 29.178, Mean Absolute Percentage Error (MAPE) sebesar 4.451, dan koefisien determinasi (R^2) sebesar 0.753. Namun, hasil MAE dengan penerapan regresi masih dapat ditingkatkan.

Weber, V. de M., et al. (2020), melakukan estimasi bobot sapi berdasarkan pengukuran dari gambar area sapi menggunakan algoritma *Random Forests* memberikan performa terbaik dengan Mean Absolute Error (MAE) sebesar 13,44 kg dan koefisien korelasi sebesar 0,75. Namun, hasil nilai MAE dengan penerapan algoritma RFR dan data yang digunakan masih dapat ditingkatkan.

Diego *et al.* (2021), mengestimasi bobot hewan domba berdasarkan gambar dengan menggunakan *machine learning regression algorithms*. Dari hasil eksperimen didapatkan metode *random forest regressor* (RFR) menghasilkan nilai error yang lebih baik dengan Mean Absolute Error (MAE) sebesar 3,099 (kg) dibandingkan dengan metode-metode *machine learning regression algorithms* yang lain. Namun, hasil nilai MAE dengan penerapan *machine learning regression algorithm* metode RFR masih dapat ditingkatkan.

Penelitian lainnya tentang estimasi bobot babi dalam menentukan kondisi kesehatan babi (Yang Chan et al., 2019). Penelitian ini menggunakan algoritma *Faster-RCNN* yang menghasilkan rata-rata error MAE sebesar 0,644 kg dari bobot aslinya. Namun, postur babi yang berbeda akan mempengaruhi nilai akurasi bobot babi. Dengan menambah data latih diharapkan akurasi dapat ditingkatkan.

Pentingnya performa dan manajemen ternak untuk perhitungan bobot domba (Zaenab et al., 2023), pada penelitiannya menerapkan metode multiple linear regression. Dari hasil penelitian didapatkan nilai MAE terkecil sebesar 0,883 dan akurasi sebesar 77,40%. Dengan melakukan kombinasi variabel berdasarkan nilai koefisien pada model persamaan regresi, *body length* (BL), *chest girth* (CG), and *shoulder height* (SH) dihasilkan nilai *body weight* (BW) = $67,004 + 0.589BL + 0.453CG + 0.399SH$. Namun, untuk nilai kombinasi yang lainnya masih bisa ditingkatkan.

Penelitian ini mengestimasi bobot sapi menggunakan 2 model machine learning. Kedua model yang diteliti dalam penelitian ini yaitu *linear regression* dan *random forest regressor* (RFR). Pentingnya mengevaluasi kinerja model menggunakan k-fold cross validation dapat menyeimbangkan antara efisiensi komputasi dan estimasi kinerja yang andal (Hastie et al., 2009).

Keuntungan penggunaan RFR, model ini mampu menggeneralisasikan data dengan kinerja yang baik (Diego et al., 2021). Memiliki kemampuan untuk menghasilkan model yang akurat dalam melakukan klasifikasi dan regresi (Breiman, L., 2001). *Random Forests* dapat mengurangi overfitting model, waktu komputasi yang efisien, menyeimbangkan bobot data, dan pemilihan fitur yang relevan (Breiman, L., 2001). Pentingnya estimasi dalam pemilihan fitur yang paling relevan untuk meningkatkan kinerja model estimasi bobot sapi yang lebih akurat (Weber, V. de M., et al., 2020). RFR mampu mengolah data berukuran besar dengan waktu komputasi yang efisien (Cabaneros, S.M., et al., 2019). *Random Forests* memiliki ketahanan terhadap outlier noise, menangani data

berdimensi tinggi secara efektif, menangkap hubungan non-linier, dan menyediakan estimasi kepentingan fitur (Breiman, L., 2001). RFR mampu memberikan hasil estimasi bobot sapi yang lebih baik dengan tingkat akurasi yang tinggi (Weber et al., 2020).

Sedangkan *linear regression* dapat menginterpretasikan hubungan linier antara variabel dependen dan independent (DA. Sant'Ana et al., 2021). Dalam konteks estimasi bobot sapi yang dilakukan kemampuannya memberikan interpretasi yang mudah dipahami terhadap hubungan antara variabel input dan output, serta kecepatan komputasi yang relatif lebih cepat dibandingkan dengan beberapa metode machine learning lainnya. Interpretasi dalam penelitian ini adalah untuk menjelaskan dan memahami hubungan antara variabel input dan output dengan cara yang mudah dipahami, serta memberikan insight tentang faktor-faktor yang mempengaruhi estimasi berat badan sapi Hereford. Memiliki kemampuan performa dalam estimasi, analisis hubungan, pemilihan variabel, evaluasi model, dan inferensi kausal (Montgomery et al., 2012). Dalam seleksi fitur *linear regression* dapat digunakan untuk pemilihan fitur dengan menganalisis signifikansi koefisien untuk menguji asumsi linearitas, independensi, homoskedastisitas, dan normalitas residu (DA. Sant'Ana et al., 2021). Sebagai model algoritma *machine learning* yang lebih kompleks dan memberikan tolak ukur yang dapat ditafsirkan untuk membandingkan kinerja model lain (DA. Sant'Ana et al., 2021). Penggunaan *machine learning* dalam penelitian ini memberikan manfaat dalam meningkatkan akurasi estimasi, kemampuan adaptasi model estimasi, efisiensi waktu dan sumber daya (Huma et al., 2019). Peneliti

berharap hasil penelitian ini dapat meningkatkan hasil nilai MAE yang lebih kecil, serta dapat memberikan kontribusi terhadap penelitian dibidang *Machine Learning*.

1.2. Rumusan Masalah

Latar belakang diatas menghasilkan rumusan masalah sebagai berikut:

- a. Berapa tingkat nilai error MAE, MAPE, RMSE, dan R-Squared terkecil untuk estimasi bobot sapi menggunakan model linear regression dan random forest regressor (RFR)?
- b. Model algoritma manakah tingkat nilai error MAE, MAPE, RMSE, dan R-Squared yang terkecil?

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

- a. Melakukan estimasi bobot sapi hereford.
- b. Indikator yang digunakan untuk perbandingan hasil performa dari random forest regressor (RFR) dan linear regression menggunakan nilai error MAE, MAPE, RMSE, dan R-Squared.
- c. Jumlah total dataset 1523 data sapi hereford.
- d. Penelitian ini tidak mengolah gambar.
- e. Penelitian ini tidak mengestimasi usia sapi.
- f. Penelitian ini tidak mengestimasi kualitas daging sapi.
- g. Penelitian ini menggunakan dataset sapi hereford.

- h. Menggunakan 13 variabel yaitu berat badan hidup, umur sapi, tinggi layu, tinggi sakrum, kedalaman dada, lebar dada, lebar maclock, lebar sendi pinggul, panjang badan miring, panjang belakang miring, lingkaran dada, ketebalan metakarpus, dan setengah ketebalan bagian belakang.
- i. Penelitian ini menggunakan nilai 10-Fold Cross Validation.
- j. Penelitian ini menggunakan 9 Feature Selection terbaik.
- k. Platform penelitian menggunakan Google Colaboratory.
- l. Penelitian ini tidak menghitung mean, standar deviasi (SD), dan koefisien variasi (CV) untuk masing-masing parameter.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

- a. Mengetahui model yang tepat untuk estimasi bobot sapi hereford.
- b. Mengetahui model yang menghasilkan nilai error MAE, MAPE, RMSE, dan R-Squared yang terkecil.

1.5. Manfaat Penelitian

Manfaat penelitian ini adalah:

- a. Bagi peternak sapi dapat mengetahui estimasi bobot sapi hereford secara akurat dan efisien.
- b. Dapat menjadi kontribusi ilmiah dalam peningkatan performa estimasi bobot sapi menggunakan *random forest regressor* dan *linear regression*.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Penelitian terdahulu yang pernah dilakukan, relevan dan dijadikan studi literatur adalah sebagai berikut:

Hasil penelitian (A N Ruchay., et al 2022), dengan menggunakan algoritma pembelajaran mesin terbaru, khususnya ExtraTreesRegressor, penelitian ini berhasil mengungkap bahwa metode ini mampu memberikan estimasi berat badan sapi Hereford yang paling akurat. Hasilnya, diukur dengan menggunakan beberapa metrik evaluasi kinerja, menunjukkan tingkat akurasi yang luar biasa, dengan Mean Absolute Error (MAE) sebesar 23.154, Root Mean Squared Error (RMSE) sebesar 29.178, Mean Absolute Percentage Error (MAPE) sebesar 4.451, dan koefisien determinasi (R^2) sebesar 0.753. Penelitian ini dilakukan dengan mempertimbangkan berbagai ukuran morfologis dan usia sapi, menggambarkan kesempurnaan algoritma ini dalam menangani keragaman data dari berbagai aspek.

Berdasarkan hasil penelitian (A N Ruchay., et al 2021), menunjukkan bahwa algoritma pembelajaran mesin, khususnya RandomForestRegressor, memberikan hasil terbaik dalam estimasi berat badan sapi Hereford nilai MAE sebesar 24.965, RMSE sebesar 37.259, MAPE sebesar 4.853 dan R^2 sebesar 0.644 berdasarkan berbagai ukuran morfologis dan usia. Namun, hasil nilai MAE dengan penerapan algoritma *machine learning* masih bisa ditingkatkan.

Berdasarkan penelitian yang dilakukan oleh (Diego et al., 2021) menggunakan learning regression algorithms untuk melakukan pengujian terhadap kelayakan dari sistem yang dibuat untuk estimasi bobot domba. Penelitian ini menggunakan teknik Random Forest Regressor menghasilkan performa terbaik dengan MAE sebesar 3,099 ($\pm 1,52$) kg. Peneliti melakukan eksperiment 80% data training dan 20% data testing menggunakan Analisis Varians (ANOVA). Sehingga masih dapat ditingkatkan mengenai data latih dan nilai akurasi.

Penelitian lain yang dilakukan oleh (Huma at al., 2019) machine learning memiliki potensi untuk estimasi berat badan ternak dengan akurat berdasarkan ukuran biometrik. Peneliti menggunakan metode random forest menghasilkan nilai koefisien determinasi sebesar 0.896 hingga 0.988 dan nilai MAE 1.242. Metode ini memberikan hasil yang akurat dan stabil, serta dapat digunakan untuk mengembangkan model estimasi berat badan dengan tingkat akurasi yang tinggi. Namun, untuk ukuran sampel dan variabel inputnya masih bisa ditambah untuk meningkatkan keakuratan dan generalisasi model estimasi.

Penelitian oleh (Alexey Ruchay at al., 2022) menggunakan dataset latih (70%), dataset uji (30%), dan dataset validasi (20% dari dataset latih). Penelitian ini menggunakan algoritma Stacking Regressor menghasilkan performa terbaik dalam estimasi bobot babi dengan MAE sebesar 4.331 dan MAPE sebesar 4.296 pada dataset uji. Peneliti menggunakan dataset 340 babi dan model yang diusulkan dapat estimasi berat babi dalam rentang 86 hingga 113 kg. Namun, penambahan fitur seperti jenis pakan, lingkungan pemeliharaan, dan faktor genetik dapat

membantu meningkatkan akurasi estimasi bobot babi.

Penelitian berikutnya yang dilakukan oleh (Akhmad Dakhlan et al., 2020) menggunakan analisis regresi. Peneliti melakukan eksperiment menghasilkan nilai korelasi antara pengukuran tubuh dan berat badan ternak. Besaran nilainya sebagai berikut body length (BL) 0.744, chest girth (CG) 0.838, shoulder height (SH) 0.543. Menghasilkan model regresi, chest girth (CG) menjadi prediktor terbaik untuk body weight (BW) dengan nilai R^2 tertinggi (0.76) dan R^2 terendah (2.795). Nilai (Akaike Information Criterion) AIC adalah 205.51 dan (Bayesian Information Criterion) BIC adalah 212.36. Namun hubungan korelasi antara ukuran tubuh dan berat badan masih bisa ditingkatkan dengan melakukan pengukuran berulang pada periode waktu yang berbeda.

Penelitian oleh (Alek Ibrahim et al., 2021) bertujuan untuk mengidentifikasi prediktor terbaik antara pengukuran tubuh dan berat badan ternak. Peneliti menggunakan metode analisis regresi linier sederhana, regresi linier berganda, dan regresi linier otomatis. Menghasilkan nilai besaran korelasi tertinggi body weight (BW) 0.866, body length (BL) 0.433 dan shoulder height (SH) 0.369 kombinasi prediktor terbaik yaitu panjang tubuh dan lingkar dada. Rumus regresi terbaik untuk estimasi berat badan adalah $BW = -56.522 + 0.509BL + 0.843CG$. Namun, dataset yang digunakan perlu lebih jelas dan terperinci. Sehingga dapat ditingkatkan mengenai data sampel yang lebih besar dan representatif.

2.2. Keaslian Penelitian

Tabel 2.1. Matriks Literatur Review dan Posisi Penelitian
Estimasi Bobot Sapi Menggunakan Linear Regression dan Random Forest Regressor

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	<i>Comparative analysis of machine learning algorithms for predicting live weight of Hereford cows</i>	Ruchay, A., Kober, V., Dorofeev, K., Kolpakov, V., Dzhulamanov, K., Kalschikov, V., & Guo, H., Computers and Electronics in Agriculture, 2022	Penelitian ini bertujuan untuk estimasi bobot sapi menggunakan metode regression machine learning.	Algoritma yang digunakan adalah algoritma regresi dengan metode ExtraTreesRegressor Mean Absolute Error (MAE) sebesar 23.154, Root Mean Squared Error (RMSE) sebesar 29.178, Mean Absolute Percentage Error (MAPE) sebesar 4.451, dan koefisien determinasi (R2) sebesar 0.753.	Penelitian kedepannya diharapkan dapat mengembangkan perbandingan algoritma yang lain dan menambah k-fold cross validation yang lebih besar untuk meningkatkan keakuratan model estimasi.	Melakukan perbandingan algoritma dengan menggunakan random forest regressor dan linear regression untuk mengetahui nilai terbaik dari setiap model dalam melakukan estimasi bobot sapi.
2	<i>Predicting the body weight of Hereford cows using machine learning</i>	Ruchay, A. N., Kolpakov, V. I., Kalschikov, V. V., Dzhulamanov, K. M., & Dorofeev, K. A., IOP, 2021	Tujuan penelitian ini adalah untuk membandingkan model regresi linear tradisional dengan algoritma pembelajaran mesin dalam estimasi bobot sapi Hereford.	Metode random forests menghasilkan performa terbaik dengan hasil nilai MAE sebesar 24.965, RMSE sebesar 37.259, MAPE sebesar 4.853 dan R2 sebesar 0.644 dalam estimasi bobot sapi Hereford.	Penelitian kedepannya diharapkan dapat mengembangkan perbandingan algoritma yang lain dan menambah kumpulan data yang lebih besar untuk meningkatkan keakuratan model estimasi.	Melakukan estimasi bobot sapi menggunakan metode random forest regressor dan linear regression untuk membandingkan tingkat keakuratan estimasi.

Tabel 2.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	<i>Weighing live sheep using computer vision techniques and regression machine learning</i>	Sant'Ana, D. A., Pacheco, M. C. B., Martins, J., Soares, W. P., de Melo, S. L. N., Garcia, V., ... & Pistori, H., Elsevier, 2021	Penelitian ini bertujuan untuk estimasi berat badan domba menggunakan metode regression machine learning.	Model Random Forest Regressor mampu menghasilkan estimasi berat badan domba dengan tingkat kesalahan yang rendah, dengan MAE sebesar 3.099 kg dan Adjusted R ² sebesar 0.687.	Penelitian ini hanya menggunakan 32 ekor domba, yang mungkin tidak mewakili keseluruhan populasi. Sehingga masih dapat ditingkatkan mengenai data latih dan nilai akurasi.	Menerapkan model random forest regressor dan linear regression untuk estimasi bobot ternak.
4	<i>A Comparative Study of Machine Learning Methods for Predicting Live Weight of Duroc, Landrace, and Yorkshire Pigs</i>	Ruchay, A., Gritsenko, S., Emolova, E., Bochkarev, A., Ermolov, S., Guo, H., & Pezzuolo, A., Animals, 2022	Tujuan penelitian ini adalah estimasi bobot babi menggunakan metode machine learning.	Algoritma Stacking Regressor memberikan performa terbaik dalam estimasi bobot babi menghasilkan mean absolute error (MAE) sebesar 4.331 dan mean absolute percentage error (MAPE) sebesar 4.296.	Hasil penelitian menunjukkan bahwa model algoritma stacking regressor memberikan kinerja estimasi terbaik. Namun, perlu menambah ukuran sampel dan memasukkan lebih banyak jenis ras babi untuk meningkatkan akurasi estimasi babi.	Penulis melakukan estimasi bobot sapi menggunakan dataset sapi untuk membandingkan metode random forest regressor dan linear regression.

Tabel 2.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
5	<i>Regression Models and Correlation Analysis for Predicting Body Weight of Female Ettawa Grade Goat using its Body Measurements</i>	Dakhlan, A., Hamdani, M., & Sulastri, S., <i>Advances in Animal and Veterinary Sciences</i> , 2020	Tujuan dari penelitian ini adalah untuk menganalisis korelasi dan model regresi terbaik untuk estimasi bobot kambing betina menggunakan pengukuran tubuh.	Besaran nilainya sebagai berikut body length (BL) 0.744, chest girth (CG) 0.838, shoulder height (SH) 0.543. Menghasilkan model regresi, chest girth (CG) menjadi prediktor terbaik untuk body weight (BW) dengan nilai R^2 tertinggi (0.76) dan R^2 terendah (2.795). Nilai (Akaike Information Criterion) AIC adalah 205.51 dan (Bayesian Information Criterion) BIC adalah 212.36.	Penelitian ini menunjukkan bahwa pengukuran chest girth (CG) dan body length (BL) dapat digunakan sebagai prediktor yang baik untuk estimasi bobot kambing, namun hubungan korelasi antara ukuran tubuh dan berat badan masih bisa ditingkatkan dengan melakukan pengukuran berulang pada periode waktu yang berbeda.	Penulis mengambil intisari dari penelitian sebelumnya yaitu dalam menganalisis pengukuran chest girth (CG) dan body length (BL) dalam estimasi bobot sapi.
6	<i>Regression model analysis for prediction of body weight from body measurements in female Batur sheep of Banjarnegara District, Indonesia</i>	Ibrahim, A., Artama, W. T., Budisatria, I. G. S., Yuniawan, R., Atnoko, B. A., & Widayanti, R., <i>Biodiversitas Journal of Biological Diversity</i> , 2021	Tujuan dari penelitian ini adalah untuk mengestimasi korelasi antara pengukuran tubuh (panjang tubuh, lingkaran dada, dan tinggi pundak) dengan bobot pada domba Batur betina.	Menghasilkan nilai besaran korelasi tertinggi body weight (BW) 0.866, body length (BL) 0.433 dan shoulder height (SH) 0.369 kombinasi prediktor terbaik yaitu panjang tubuh dan lingkaran dada. Rumus regresi terbaik untuk mengestimasi berat badan adalah $BW = -56.522 + 0.509BL + 0.843CG$.	Dataset yang digunakan perlu lebih jelas dan terperinci. Sehingga dapat ditingkatkan mengenai data sampel yang lebih besar dan representatif.	Penulis mengambil intisari dari penelitian sebelumnya yaitu dalam analisis metode regresi linier. Perbedaan terletak pada tujuan penelitian serta metode yang digunakan pada penelitian.

2.3. Landasan Teori

2.3.1. Bobot Sapi

Bobot sapi merupakan sifat ekonomi yang membantu dalam pemilihan hewan untuk ditanakkan (J. Patel Ashwini et al., 2019). Menjadi salah satu parameter penting dalam manajemen peternakan sapi, karena dapat digunakan untuk mengukur pertumbuhan dan perkembangan sapi, menentukan kebutuhan pakan, serta mengestimasi produksi daging dan reproduksi sapi (Weber et al., 2020). Ukuran numerik yang menggambarkan berat badan atau massa tubuh sapi sebagai indikator penting dalam manajemen ternak, pemilihan genetik, perencanaan pakan, dan evaluasi kesehatan dan kinerja sapi (Roche et al., 2009). Bobot sapi sebanding dengan bobot, usia, dan masa menyusunya guna menghasilkan susu yang berkualitas (Kanuya et al., 2006). Perkembangan peternakan sapi dapat memastikan bahwa produksi hewan ternak dapat memenuhi kebutuhan manusia secara berkelanjutan dan etis, sambil menjaga kesejahteraan hewan dan menjaga lingkungan alam sekitar. Peternakan sapi sering menjadi sumber pendapatan utama bagi masyarakat di pedesaan. Dapat memberikan lapangan kerja bagi banyak orang, baik dalam hal pemeliharaan sapi, pengolahan produk susu dan daging, maupun dalam distribusi dan pemasaran produk-produk tersebut (William, 2021). Peternakan sapi juga memberikan kesempatan bagi petani dan peternak untuk menghasilkan pendapatan dan meningkatkan kesejahteraan. Bobot sapi memiliki potensi untuk inovasi dan pengembangan lebih lanjut untuk penggunaan teknologi seperti pemantauan Kesehatan, reproduksi, dan penggunaan pakan yang lebih efisien (William, 2021). Pentingnya pengukuran tubuh dapat

digunakan untuk estimasi bobot sapi. Ada korelasi yang erat antara berat badan dan ukuran tubuh (Ozkaya dan Bozkurt, 2009). Seperti yang dapat dilihat pada Gambar 2.1.



Gambar 2.1. Hewan Ternak Sapi

Sumber: (Kuhl et al., 2019)

2.3.2. Machine Learning

Machine learning masuk ke dalam subbidang dari artificial intelligence (AI) yang didefinisikan sebagai kemampuan sebuah mesin untuk mempelajari kecerdasan manusia (Geron, 2017). Subbidang kecerdasan buatan yang berfokus pada pengembangan algoritma dan model yang memungkinkan komputer mempelajari dan membuat estimasi atau keputusan tanpa diprogram secara eksplisit (Hierons et al., 1999). Ini melibatkan penggunaan teknik statistik dan algoritma untuk menganalisis dan menafsirkan data dalam jumlah besar, memungkinkan komputer mengidentifikasi pola, membuat estimasi, dan meningkatkan kinerjanya seiring waktu (Hierons et al., 1999). Berikut adalah beberapa manfaat utama menggunakan machine learning:

a. Pengambilan Keputusan yang Lebih Baik

Algoritma machine learning dapat menganalisis data dalam jumlah besar dan mengidentifikasi pola dan tren yang mungkin tidak terlihat oleh manusia. Hal ini memungkinkan organisasi untuk membuat keputusan yang lebih tepat dan

berdasarkan data (Davenport et al., 2018).

b. Otomatisasi dan Efisiensi

Pembelajaran mesin dapat mengotomatiskan tugas dan proses yang berulang, menghemat waktu dan sumber daya. Hal ini juga dapat meningkatkan efisiensi dengan mengoptimalkan proses dan alur kerja yang kompleks (Chen et al., 2014).

c. Sistem Personalisasi dan Rekomendasi

Algoritma pembelajaran mesin dapat menganalisis data dan perilaku pengguna untuk memberikan rekomendasi dan pengalaman yang dipersonalisasi. Ini banyak digunakan di platform e-commerce, hiburan, dan konten (Ricci et al., 2011).

d. Deteksi Penipuan dan Keamanan Siber

Pembelajaran mesin dapat mendeteksi pola dan anomali dalam data, sehingga efektif dalam deteksi penipuan dan keamanan siber. Ini dapat mengidentifikasi transaksi penipuan, mendeteksi intrusi jaringan, dan melindungi informasi sensitive (Bhattacharyya et al., 2014).

e. Perawatan Kesehatan dan Diagnosis Medis

Algoritma pembelajaran mesin dapat menganalisis data medis dan membantu mendiagnosis penyakit, estimasi hasil pasien, dan merekomendasikan rencana perawatan. Hal ini dapat meningkatkan akurasi dan efisiensi pemberian layanan Kesehatan (S. Kumar et al., 2017).

f. Pemrosesan Bahasa Alami

Pembelajaran mesin memungkinkan komputer memahami dan memproses

bahasa manusia. Ini digunakan dalam asisten virtual, chatbots, dan aplikasi terjemahan Bahasa (S. Kumar et al., 2017).

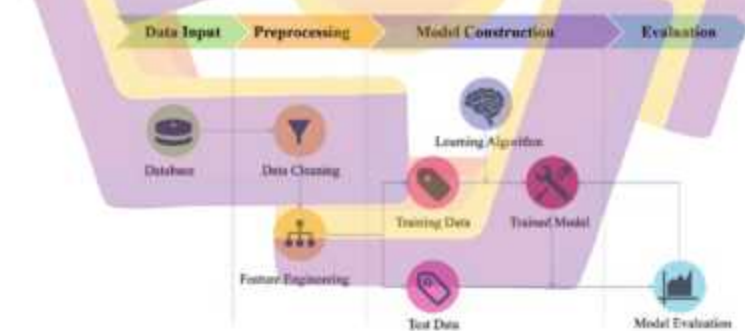
g. Pengenalan Gambar dan Ucapan

Algoritma pembelajaran mesin dapat menganalisis dan menafsirkan gambar dan ucapan, memungkinkan aplikasi seperti pengenalan wajah, deteksi objek, dan konversi ucapan menjadi teks (A. Kimura et al., 2004).

h. Analisis Prediktif

Pembelajaran mesin dapat menganalisis data historis untuk membuat estimasi dan prakiraan. Ini digunakan di berbagai domain, termasuk keuangan, pemasaran, dan manajemen rantai pasokan (A. Tharwat et al., 2014).

Untuk memahami struktur data dan menyesuaikan data tersebut kedalam sebuah model untuk melakukan berbagai tugas kompleks. Seperti yang terlihat pada Gambar 2.2.



Gambar 2.2. Alur Kerja Machine Learning

Sumber: (Yang et al., 2022)

Secara umum terdapat beberapa skenario yang digunakan dalam machine learning, diantaranya adalah sebagai berikut:

a. Supervised Learning

Dalam Supervised Learning komputer diberikan contoh input yang diberi label sesuai dengan output yang diinginkan. Tujuan dari digunakannya metode ini adalah agar algoritma yang digunakan dapat mempelajari data dengan membandingkan output aktual yang telah dipelajari sebelumnya untuk menemukan kesalahan dan memodifikasinya sesuai dengan model yang digunakan.

b. Unsupervised Learning

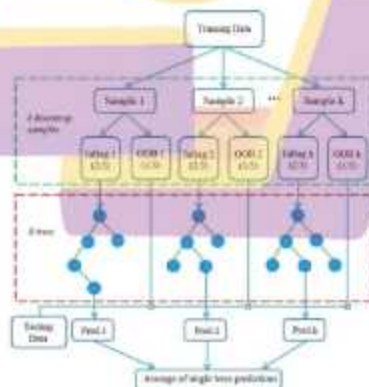
Pada Unsupervised Learning dilakukan pengolahan data yang tidak memiliki label. Algoritma yang digunakan akan mempelajari data yang diinputkan untuk mengelompokkan data yang memiliki karakteristik yang sama. Sederhananya tujuan utama dari Unsupervised Learning adalah untuk menemukan pola tersembunyi pada data yang diinputkan untuk melakukan pembelajaran terhadap model yang digunakan.

2.3.3. Random Forest Regressor

Algoritma Random Forest Regressor merupakan sebuah metode pembelajaran regresi bersifat Ensemble yang diusulkan oleh Leo Breiman (Breiman, 2001). Ensemble sendiri merupakan model yang terdiri dari beberapa algoritma, sehingga dalam Random Forest ini terkandung beberapa decision tree dengan distribusi yang sama dan dengan kondisi tidak berkorelasi satu sama lain untuk membangun forest atau "hutan" dengan tujuan melatih dan estimasi data sampel (Zhang et al., 2021). Random Forest Regression dapat mengandung ratusan atau bahkan ribuan Decision Tree yang bertindak sebagai fungsi regresi sendiri.

Setiap tree di RFR ditanam dengan subset prediktor acak maka dari itu algoritma ini disebut hutan 'acak' (Zhang et al., 2021).

Decision Tree atau yang disebut juga sebagai Classification and Regression Tree (CART) adalah sebuah model statistik nonparametrik yang dapat menggambarkan hubungan antara variabel. Variabel di sini merupakan variabel respon (dependen) dengan satu atau lebih variabel prediktor (independen). Setiap pohon keputusan terdiri dari simpul keputusan atau internal node dan simpul daun atau leaf. Setiap node keputusan mengevaluasi setiap sampel variabel atau atribut dan setiap cabangnya merupakan hasil dari pengujian tersebut, sementara itu node terluar yaitu daun menjadi labelnya. Berbeda dari hasil akhir Random Forest Classifier yang mengambil Majority Votes dari berbagai Decision Tree sebagai hasilnya, hasil akhir dari Random Forest Regression ini merupakan nilai rata-rata dari keluaran semua Decision Tree yang telah dibangun (Li et al., 2018). Gambar 2.4 akan menunjukkan cara kerja dari Random Forest Regression.



Gambar 2.4. Cara Kerja Random Forest Regressor

Sumber: (Zhang et al., 2021)

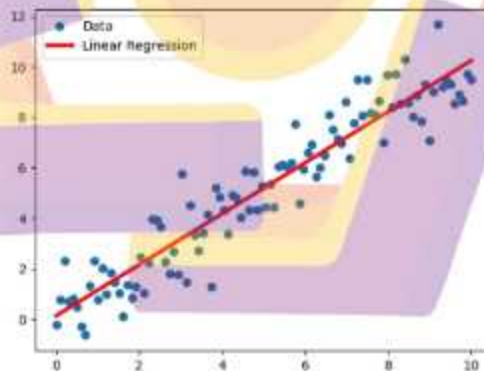
2.3.4. Linear Regression

Linear Regression merupakan salah satu jenis regresi yang melibatkan hubungan antara satu variabel dependen atau variabel tak bebas atau label (Y) dan variabel independen atau variabel bebas atau fitur (X). Besarnya nilai variabel dependen akan dipengaruhi oleh besar kecilnya variabel bebas. Algoritma Linear Regression didasarkan pada pola hubungan data terkait masa lalu (Hamdanah & Fitriana, 2021). Tujuan dari linear regression adalah untuk menemukan garis regresi yang terbaik yang dapat digunakan untuk estimasi nilai variabel dependen berdasarkan nilai variabel prediktor. Metode ini didasarkan pada asumsi bahwa hubungan antara variabel independen dan variabel dependen dapat dijelaskan dengan persamaan garis lurus untuk analisis estimasi dan juga untuk memahami hubungan kausal antara variabel-variabel tersebut (Gujarati et al., 2009). Umumnya, algoritma Linear Regression dibagi menjadi dua jenis, yaitu Simple Linear Regression dan Multiple Linear Regression. Simple Linear Regression merupakan hubungan antara satu variabel dependen dengan satu variabel independen, sedangkan Multiple Linear Regression merupakan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen (Herwanto et al., 2019). Pada penelitian ini, akan digunakan Multiple Linear Regression yang diekspresikan melalui Persamaan 2.1.

$$Y = a + a_1X_1 + b_2X_2 + \dots + b_nX_n \quad (2.1)$$

Terdapat beberapa komponen pada persamaan 2.1, di antaranya adalah Y yang merupakan variabel dependen atau nilai yang diestimasi, a adalah konstanta, X_n adalah variabel independen, dan b_n adalah koefisien regresi. Dari

persamaan ini, dapat ditarik sebuah garis yang mampu mengestimasi variabel dependen berdasarkan variabel independen. Algoritma Multiple Linear Regression akan berusaha untuk menemukan garis estimasi terbaik (Gupta et al., 2020). Hasil dari linear regression dapat diinterpretasikan dengan mudah. Koefisien regresi dapat memberikan informasi tentang sejauh mana variabel prediktor mempengaruhi variabel dependen (Gujarati et al., 2009). Kualitas garis estimasi dapat ditentukan dari seberapa dekatnya garis estimasi dengan poin-poin data nilai variabel dependen. Linear regression dapat digunakan untuk mengestimasi nilai variabel dependen berdasarkan nilai variabel prediktor. Jika asumsi linearitas terpenuhi, model regresi linier dapat memberikan estimasi yang akurat (Gujarati et al., 2009). Berikut Gambar 2.5 yang menggambarkan bagaimana bentuk garis Linear Regression.



Gambar 2.5. Garis Linear Regression

Sumber: (Tran, 2019)

Pada Gambar 2.5 garis yang berwarna merah merupakan garis Linear Regression dan poin-poin berbentuk bulat berwarna biru merupakan poin data dari variabel independen. Sebuah garis Linear Regression dapat dikatakan bagus apabila garis mempunyai jarak yang dekat dengan keseluruhan poin-poin data variabel independen. Semakin dekat garis Linear Regression dengan poin-poin data variabel independen, maka akan semakin bagus juga estimasi yang dihasilkan.

2.3.5. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) merupakan salah satu pengujian dari model regresi. MAE ini menghitung rata-rata perbedaan mutlak antara nilai yang sebenarnya dan nilai hasil estimasi model yang dibangun. Semua kesalahan pengujian memiliki bobot yang sama pada MAE. Semakin kecil nilai MAE maka semakin akurat hasil estimasinya (Li et al., 2018). Untuk menghitung nilai MAE pada hasil pengujian model regresi, digunakan Persamaan 2.2 berikut. Pada persamaan 2.2, n merepresentasikan jumlah observasi yang dilakukan, y_i merepresentasikan experimental values, dan $\hat{y}_i$ merepresenmtasikan nilai hasil estimasi model.

$$MAE = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \quad (2.2)$$

Mean Square Error (MSE) merupakan nilai kuadrat dari perbandingan kualitas kesesuaian antara data aktual dan data hasil model estimasi (Islam, 2021). Namun terkadang jika model regresi pada akhirnya menghasilkan suatu estimasi yang sangat buruk, bagian kuadrat dari fungsi MSE dapat memperbesar kesalahan sehingga Root Mean Square Error (RMSE) dapat menjadi solusi. MAE memberikan gambaran tentang sejauh

mana estimasi model berbeda dari nilai aktual. Manfaat utama dari MAE adalah sebagai berikut (Weber et al., 2020):

a. Mengukur kesalahan secara absolut

Mengukur kesalahan estimasi secara absolut tanpa memperhatikan arah kesalahan. Ini berarti bahwa MAE memberikan informasi tentang sejauh mana estimasi model berbeda dari nilai aktual tanpa memperhatikan apakah estimasi tersebut terlalu tinggi atau terlalu rendah.

b. Interpretasi yang mudah

MAE diukur dalam satuan yang sama dengan variabel target, sehingga lebih mudah untuk diinterpretasikan. Misalnya, jika kita estimasi berat sapi dalam kilogram, MAE juga akan diukur dalam kilogram.

c. Robust terhadap outlier

MAE tidak terlalu sensitif terhadap adanya outlier dalam data. Hal ini karena MAE menggunakan nilai absolut dari kesalahan, sehingga satu atau beberapa outlier tidak akan memiliki pengaruh yang besar pada nilai keseluruhan MAE.

d. Digunakan dalam pemilihan model

MAE dapat digunakan untuk membandingkan performa model yang berbeda. Model dengan MAE yang lebih rendah dianggap memiliki performa yang lebih baik dalam mengestimasi nilai aktual.

2.3.6. Root Mean Square Error (RMSE)

Nilai dari Root Mean Square Error (RMSE) adalah hasil dari akar kuadrat hasil Mean Square Error (MSE) (Islam, 2021). Root Mean Square Error (RMSE) sendiri merupakan salah satu kriteria yang paling umum digunakan untuk model regresi (Rachman, 2018). MSE dan RMSE ini mengukur besarnya tingkat error atau kesalahan

yang terjadi dari hasil estimasi oleh model, dimana semakin kecil nilai yang didapatkan (mendekati 0), maka hasil estimasi akan semakin akurat (Suprayogi et al., 2014). Nilai dari MSE dan RMSE ini dapat dihitung dengan menggunakan Persamaan 2.3 dan 2.4.

$$MSE = \frac{\sum_{t=1}^n (Y_t - Y'_t)^2}{n} \quad (2.3)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - Y'_t)^2}{n}} \quad (2.4)$$

Pada Persamaan 2.3 dan 2.4 di atas, Y_t merupakan nilai aktual atau sebenarnya pada periode t , Y'_t adalah nilai hasil estimasi pada periode t , dan n menunjukkan jumlah estimasi yang dilakukan. Manfaat dari RMSE adalah sebagai berikut (Diego et al., 2021):

a. Mengukur kesalahan estimasi secara keseluruhan

RMSE memberikan gambaran tentang sejauh mana model regresi dapat estimasi nilai dengan akurat secara keseluruhan. Dengan menghitung akar kuadrat dari selisih kuadrat, RMSE memberikan bobot yang lebih besar pada kesalahan yang lebih besar, sehingga memberikan gambaran yang lebih baik tentang kualitas estimasi model.

b. Membandingkan performa model

RMSE memungkinkan perbandingan langsung antara model regresi yang berbeda. Semakin kecil nilai RMSE, semakin baik performa model dalam estimasi nilai sebenarnya. Dengan membandingkan RMSE dari beberapa model, kita dapat memilih model yang memiliki performa terbaik dalam estimasi data.

c. Interpretasi yang mudah

RMSE memiliki satuan yang sama dengan variabel target, sehingga lebih mudah untuk diinterpretasikan. Misalnya, jika kita estimasi berat domba dalam

kilogram, RMSE juga akan dalam satuan kilogram. Hal ini memudahkan dalam memahami sejauh mana kesalahan estimasi dalam konteks yang relevan.

d. Sensitivitas terhadap outlier

RMSE lebih sensitif terhadap outlier metrik evaluasi lainnya seperti Mean Absolute Error (MAE). Karena RMSE mengkuadratkan selisih antara nilai estimasi dan nilai sebenarnya, outlier akan memiliki dampak yang lebih besar pada nilai RMSE. Hal ini dapat membantu dalam mengidentifikasi dan menangani outlier yang mungkin mempengaruhi performa model.

2.3.7. R square (R²)

R square yang biasanya disimbolkan dengan R² bertujuan untuk mengukur seberapa jauh kemampuan model regresi yang dibangun dalam menerangkan pengaruh dari variasi variabel dependen (tidak bebas) (Manurung, 2015). Uji ini dinamakan koefisien determinasi karena variasi yang terjadi dalam variabel tak bebas (Y) dapat dijelaskan oleh variabel bebas (X) dengan adanya regresi linier Y atas X. Besar nilai dari koefisien determinasi adalah berkisar $0 \leq R^2 \leq 1$. Jika R² mendekati 1 maka dapat dikatakan pengaruh variabel bebas terhadap variabel terikat adalah besar yang berarti model yang digunakan baik untuk menjelaskan pengaruh variabel tersebut (Harahap et al., 2013). Nilai dari R² dapat ditemukan menggunakan persamaan 2.5.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (2.5)$$

Ukuran statistik yang digunakan untuk mengukur sejauh mana variasi variabel dependen dapat dijelaskan oleh variabel independen dalam model regresi.

R² menggambarkan proporsi variasi dalam variabel dependen yang dapat dijelaskan oleh variabel independen dalam model regresi. Nilai R² berkisar antara 0 hingga 1, di mana nilai 0 menunjukkan bahwa variabel independen tidak dapat menjelaskan variasi dalam variabel dependen, sedangkan nilai 1 menunjukkan bahwa variabel independen dapat menjelaskan seluruh variasi dalam variabel dependen. R² dapat diinterpretasikan sebagai persentase variasi dalam variabel dependen yang dapat dijelaskan oleh variabel independen dalam model regresi (Diego et al., 2021).

2.3.8. Mean Absolute Percentage Error (MAPE)

MAPE dihitung dengan menggunakan kesalahan absolut pada tiap periode dibagi dengan nilai observasi yang nyata untuk periode itu. Kemudian, merata-rata kesalahan persentase absolut tersebut. Pendekatan ini berguna ketika ukuran atau besar variabel estimasi itu penting dalam mengevaluasi ketepatan estimasi. MAPE mengindikasikan seberapa besar kesalahan dalam estimasi yang dibandingkan dengan nilai nyata (Siami et al., 2018). Tujuan Mean Absolute Percentage Error (MAPE) adalah untuk mengukur tingkat kesalahan atau deviasi antara nilai estimasi dan nilai aktual dalam bentuk persentase untuk mengevaluasi kualitas estimasi (Hyndman et al., 2006). Nilai dari MAPE dapat ditemukan menggunakan persamaan 2.6

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \quad (2.6)$$

Berikut ini adalah rumus *Linear Regression*, *Random Forest Regressor*, *Mean Absolute Error* (MAE), *Root Mean Absolute Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) dan *R-square* dibawah ini:

$$A = b + b_1 Z_1 + c_2 Z_2 + \dots + c_n Z_n \quad (2.7)$$

$$l(y) = \arg \max_c (\sum_{i=1}^Z Y_{an(y)=c}) \quad (2.8)$$

$$MAE = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \quad (2.9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (2.10)$$

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \quad (2.11)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (2.12)$$

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Jenis penelitian yang dilakukan termasuk ke dalam penelitian kuantitatif. Penelitian yang dilakukan termasuk kedalam penelitian komputasi eksperimental dengan melakukan pengujian terhadap performa estimasi bobot sapi menggunakan linear regression dan random forest regressor. Eksperimen ekspremental adalah subjek penelitian yang akan dilakukan, di mana peneliti akan melakukan sebuah eksperimen untuk menemukan estimasi berat badan sapi menggunakan linear regression dan random forest regressor. Peneliti menggunakan pendekatan kuantitatif, yang berarti bahwa penelitian akan dilakukan sesuai dengan alur yang telah dibuat.

3.2. Metode Pengumpulan Data

Pengumpulan data dilakukan untuk mendapatkan data yang diperlukan dalam penelitian ini. Dataset yang digunakan adalah dataset sapi hereford yang didapatkan dari Github. Data dari 1.523 sapi Hereford dikumpulkan dari Agustus 2008 hingga September 2020. Sapi Hereford milik dua peternakan swasta: (1) dengan pakan terkonsentrasi di wilayah Nizhny Novgorod, Rusia, dan (2) dengan pakan biji-bijian di wilayah Voronezh, Rusia. Berat sapi-sapi ini diukur menggunakan timbangan dan berkisar antara 314 hingga 750 kilogram. Umur sapi berkisar antara 3 sampai 5 tahun. Berat badan hewan juga diperkirakan

menggunakan sebelas pengukuran morfometrik yang ditunjukkan pada tongkat Lydtin yang digunakan untuk mengukur WH dan HH. Kompas Wilkens digunakan untuk mengukur CW, IW, ORL dan BHG. Terakhir, pita pengukur digunakan untuk mengukur seluruh dimensi tubuh yang tersisa (Ruchay, 2021).



Gambar 3.1. Sapi Hereford

Dapat dilihat gambar 3.1 sapi hereford merupakan salah satu jenis sapi dari bangsa sapi croa yang hidup di iklim subtropis. Karakteristik sapi hereford mempunyai warna utama pada tubuh merah dan putih. Warna putih terdapat di bagian sisi badan, muka, dada, perut bawah, bahu, ekor dan keempat kaki dari batat lututnya. Karena warna ini, sapi hereford dinamakan pula sebagai sapi bermuka putih. Tubuh sapi hereford tidak tinggi tetapi tegap dengan daging yang padat serta menampakkan urat tubuh. Garis punggung sapi hereford berbentuk rata dan lebar. Tubuh sapi hereford memanjang tetapi membulat. Lambung sapi hereford berukuran besar. Kemudian pada hari pengumpulan data, semua hewan yang dipilih dikumpulkan di pagar makanan di area terpisah untuk pengukuran tubuh manual dan perekaman gambar (Ruchay et al., 2020).

3.3. Metode Analisis Data

Setelah data yang dibutuhkan terkumpul maka, langkah selanjutnya yang akan dilakukan adalah mengolah dataset yang didapat menjadi sebuah informasi. Sebelum melakukan analisis data, data sapi yang didapatkan dilakukan pre-processing data terlebih dahulu, seperti *data labelling*, *data cleaning*, *data normalization*, dan *fitur selection*.

Untuk melakukan analisa perbandingan performa menggunakan linear regression dan random forest regressor dalam estimasi bobot sapi, dan melakukan perbandingan hasil dari kedua model tersebut. Kemudian evaluasi kinerja model menggunakan *k-fold cross validation* dan menggunakan metrik evaluasi seperti MAE, RMSE, MAPE, dan R2. Hasil pengukuran tersebut akan dijadikan acuan atau pedoman dalam menentukan hasil atau analisis estimasi bobot sapi pada penelitian ini. Platform eksperimen yang akan digunakan adalah Google Colaboratory.

3.4. Dataset

Sumber dataset dari https://github.com/ruchaya/Hereford_cows. Dataset yang digunakan pada penelitian ini adalah dataset sapi hereford menggunakan 1523 data sapi. Kemudian tahapan berikutnya yaitu preprocessing, dimana pada tahap ini dataset yang digunakan akan dibagi menjadi 13 variabel yaitu berat badan hidup, umur sapi, tinggi badan layu, tinggi sakrum, kedalaman dada, lebar dada, lebar maclock, lebar sendi panggul, panjang badan miring, panjang punggung miring, lingkaran dada, ketebalan metakarpus, dan setengah ketebalan bagian belakang.

Ada beberapa tahapan *preprocessing* yang dilakukan, dimulai dari *data labelling*, *data cleaning*, *data normalization*, dan *fitur selection*. Dataset sapi dapat dilihat di tabel 3.1 untuk proses pengolahan data *preprocessing*.

- a. *Data Labelling* adalah sebuah proses untuk memberikan klasifikasi pada setiap data sapi. Tujuannya adalah untuk mengidentifikasi dan membedakan data berdasarkan atribut tertentu. Pelabelan data dalam penelitian ini sangat penting untuk pengelompokan, pemodelan, dan analisis statistik yang lebih fokus dan relevan (Hastie et al., 2009).
- b. *Data Cleaning* adalah sebuah proses untuk menjamin kualitas data. Tujuannya adalah untuk menghilangkan data yang tidak valid, tidak lengkap, dan tidak relevan. Ini juga memberikan hasil penelitian yang akurat dan dapat diandalkan (Weber et al., 2020).
- c. *Data Normalization* adalah proses untuk mengubah data menjadi bentuk standar, sehingga memudahkan dalam mengolah dan menganalisis data ternak. Normalisasi data bertujuan untuk menghilangkan perbedaan skala untuk memastikan bahwa setiap atribut mempunyai kontribusi yang seimbang untuk memperoleh hasil penelitian yang lebih akurat (Ruchay et al., 2022).
- d. *Fitur Selection* adalah proses untuk mengidentifikasi subset fitur yang paling relevan dan signifikan dalam dataset ternak. Pemilihan fitur bertujuan untuk mereduksi dimensi data, meningkatkan efisiensi komputasi, menghilangkan komponen redundant, dan meningkatkan performa model estimasi (Weber et al., 2020).

Index	body weight	age years	withers height	height in the sacrum	chest depth	chest width	width in maclocks	hip joint width	oblique length of the body	oblique rear length	chest girth	metacarpus girth	backside half girth
0	110	0	110	108	68	47	18	28	110	90	100	30	17
1	100	1	90	100	60	50	27	25	100	80	100	30	16
2	100	1	90	100	60	50	27	25	100	80	100	30	16
3	110	0	110	108	68	47	18	28	110	90	100	30	17
4	100	1	90	100	60	50	27	25	100	80	100	30	16

Gambar 3.2. Cuplikan isi dataset sapi Hereford

Cuplikan isi dataset sapi Hereford dapat dilihat pada Gambar 3.2. pada gambar tersebut tidak dapat ditampilkan semua kolom karena keterbatasan area tulisan. Semua kolom dataset berisi angka pada setiap variabel yang digunakan seperti *live weight*, *age years*, *withers height*, *height in the sacrum*, *chest depth*, *chest width*, *width in maclocks*, *hip joint width*, *oblique length of the body*, *oblique rear length*, *chest girth*, *metacarpus girth* dan *backside half girth*. Dataset ini juga digunakan untuk estimasi bobot sapi menggunakan metode regression machine learning (Ruchay et al., 2022).

Dataset sapi Hereford memiliki manfaat pada setiap dimensi pengukuran tubuh sapi memberikan informasi yang berguna dalam mengevaluasi kondisi tubuh dan kesehatan sapi yang dapat dilihat secara keseluruhan pada Tabel 3.1. Berikut adalah manfaat dari pengukuran tubuh sapi dari setiap dimensi (Ruchay et al., 2021):

- Body Weight*: memberikan informasi tentang berat badan sapi, yang dapat digunakan untuk menentukan dosis obat yang tepat dan mengevaluasi kesehatan sapi.
- Age Years*: memberikan informasi tentang umur sapi hereford, yang dapat digunakan untuk mengevaluasi perkembangan dan kesehatan sapi.

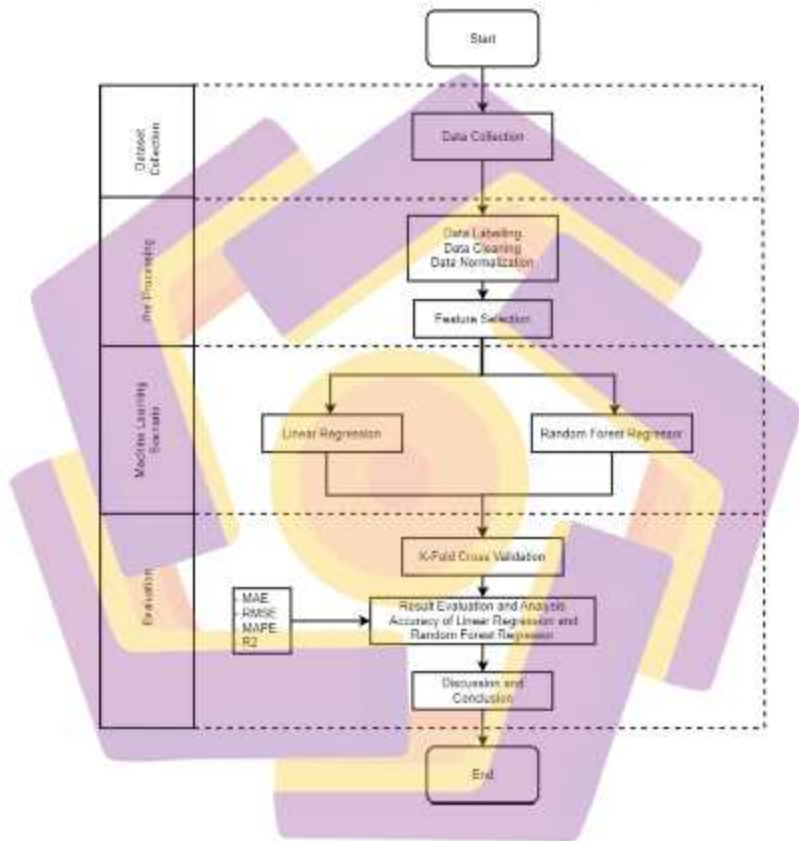
- c. *Withers height*: memberikan informasi tentang tinggi sapi pada bagian bahu, yang dapat digunakan untuk mengevaluasi pertumbuhan dan umur sapi.
- d. *Hip Height*: memberikan informasi tentang tinggi sapi pada bagian belakang, yang dapat digunakan untuk menentukan ukuran kandang yang sesuai dan mengevaluasi pertumbuhan sapi.
- e. *Chest Depth*: memberikan informasi tentang kedalaman dada sapi, yang dapat digunakan untuk mengevaluasi kondisi pernapasan sapi.
- f. *Chest Width*: memberikan informasi tentang lebar dada sapi, yang dapat digunakan untuk menentukan kesehatan dan kualitas daging sapi.
- g. *Ilium Width*: memberikan informasi tentang lebar pinggul sapi, yang dapat digunakan untuk menentukan kapasitas kandang sapi.
- h. *Sciatic Tubercle Width*: memberikan informasi tentang lebar sendi pinggul sapi, yang dapat digunakan untuk mengevaluasi kesehatan sendi sapi.
- i. *Oblique Body Length*: memberikan informasi tentang panjang tubuh sapi dari pangkal leher hingga pangkal ekor, yang dapat digunakan untuk mengevaluasi pertumbuhan dan kandang sapi yang tepat.
- j. *Oblique Rear length*: memberikan informasi tentang panjang tubuh sapi dari pangkal leher hingga pangkal ekor bagian belakang sapi.
- k. *Heart Girth*: memberikan informasi tentang lingkaran dada sapi, yang dapat digunakan untuk menentukan bobot sapi.
- l. *Metacarpus Girth*: ketebalan pada titik tertipis tulang metacarpal.
- m. *Backside Half-Girth*: jarak dari tonjolan lateral sendi lutut kiri ke belakang di bawah ekor dan ke titik yang sama pada sendi kanan.

Tabel 3.1. Cuplikan Dataset Sapi Hereford

No.	live weight	Age Years	withers height	height in the sacrum	chest depth	chest width	width in maclocks	hip joint width	oblique length of the body	oblique rear length	chest girth	metacarpus girth	backside half-girth
1.	415	5	117	122	62	40	43	42	145	43	172	20	71
2.	407	3	116	121	60	39	42	44	127	37	171	18	65
3.	448	5	114	121	60	43	41	40	128	41	176	19	68
4.	443	3	118	123	63	46	44	44	150	46	176	19	66
5.	410	3	124	127	66	41	42	44	140	45	178	20	70
6.	441	5	120	124	60	43	43	44	138	45	178	20	69
7.	427	5	117	121	62	41	43	45	149	45	172	21	70
8.	380	5	115	117	60	48	41	41	137	43	172	19	69
9.	416	5	118	120	62	44	42	40	140	44	175	21	71
10.	424	5	119	122	65	46	44	40	151	46	182	19	67
11.	450	3	120	124	64	43	43	45	138	42	180	19	66
12.	429	5	118	120	60	46	40	41	140	41	178	20	70
13.	410	5	117	119	61	41	41	40	143	44	179	21	71
14.	464	3	119	122	62	45	44	41	133	39	186	19	68
15.	407	3	116	129	64	40	39	40	139	41	175	19	68
16.	312	3	105	110	55	31	38	37	137	40	153	20	69
17.	412	5	118	120	61	43	43	45	143	45	183	20	73

3.5. Alur Penelitian

Alur penelitian diilustrasikan pada gambar 3.3. dimulai dari pengumpulan data, *preprocessing data*, skenario *machine learning*, evaluasi dan analisis hasil.



Gambar 3.3. Alur penelitian

Alur penelitian dimulai dari persiapan awal, pengumpulan data, pemrosesan data, pemodelan, dan evaluasi. Tahapan pertama pada penelitian ini adalah pengumpulan dataset yang diambil dari Github. Dataset yang digunakan pada penelitian ini adalah dataset sapi Hereford menggunakan 1523 data sapi. Kemudian

tahapan berikutnya yaitu preprocessing, dimana pada tahap ini menggunakan 10 variabel yaitu berat badan hidup, tinggi layu, tinggi sakrum, kedalaman dada, lebar dada, lebar maclock, lebar sendi pinggul, panjang badan miring, panjang belakang miring, lingkaran dada, ketebalan metakarpus, dan setengah ketebalan bagian belakang (Ruchay et al., 2021). Preprocessing dimulai dari *data labelling*, *data cleaning*, *data normalization*, dan *fitur selection*.

Kemudian *data labelling* untuk mengidentifikasi dan membedakan data berdasarkan setiap atribut seperti *live weight*, *age years*, *withers height*, *height in the sacrum*, *chest depth*, *chest width*, *width in maclocks*, *hip joint width*, *oblique length of the body*, *oblique rear length*, *chest girth*, *metacarpus girth* dan *backside half girth*. Selanjutnya proses *data cleaning* untuk menjamin kualitas data dan menghilangkan data yang tidak valid, tidak lengkap, dan tidak relevan dengan menghilangkan bagian datanya seperti *no. identificatory*.

Selanjutnya *data normalization* proses untuk mengubah data menjadi bentuk standar, sehingga memudahkan dalam mengolah dan menganalisis data sapi yang diolah menggunakan google colaboratory.

Kemudian tahapan berikutnya yaitu mempersiapkan model yang akan digunakan, dimana pada penelitian ini menggunakan linear regression dan random forest regressor untuk estimasi bobot sapi. Dalam penelitian yang menarik ini, dilakukan estimasi bobot sapi menggunakan metode regresi linear. Suatu model matematis dibangun dengan memanfaatkan variabel-variabel prediktor seperti jenis pakan (X_1), jumlah pakan yang diberikan (X_2), dan faktor-faktor lainnya (X_n). Dalam rumus regresi linear yang digunakan, Y menggambarkan bobot sapi,

sedangkan a adalah konstanta, dan $b_1, b_2, \dots, b_n$ adalah koefisien regresi yang mengukur kontribusi masing-masing variabel prediktor terhadap variabel respons. Model ini dapat memberikan wawasan yang berharga dalam upaya meningkatkan produksi sapi dan mengoptimalkan faktor-faktor yang memengaruhi pertumbuhan sapi.

Selanjutnya dengan mengadopsi metode Random Forest Regressor yang inovatif. Pendekatan ini membuka jendela baru dalam pemahaman kita terhadap faktor-faktor yang memengaruhi pertumbuhan bobot sapi. Melalui analisis yang cermat terhadap dataset yang kaya akan variabel, kami menerapkan rumus Random Forest Regressor yang kompleks namun kuat, memberikan hasil estimasi yang akurat dan dapat diandalkan. Rumusnya dapat dirumuskan sebagai berikut:

$$W = \frac{1}{N} \sum_{i=1}^N w_i \quad (3.1)$$

di mana W adalah estimasi bobot sapi, N adalah jumlah pohon dalam hutan acak, dan w_i adalah estimasi bobot dari pohon ke- i . Metode ini diharapkan dapat memberikan kontribusi signifikan dalam mendukung pengembangan teknik estimasi bobot sapi yang lebih unggul, berpotensi meningkatkan efisiensi dalam pengelolaan peternakan. Kemudian melakukan uji performa menggunakan k-fold cross validation pada setiap metode dan dataset yang digunakan.

Setelah model telah siap maka, langkah berikutnya yang harus dilakukan adalah penarikan kesimpulan dari perbandingan hasil yang telah didapatkan berdasarkan skenario yang telah dilakukan.

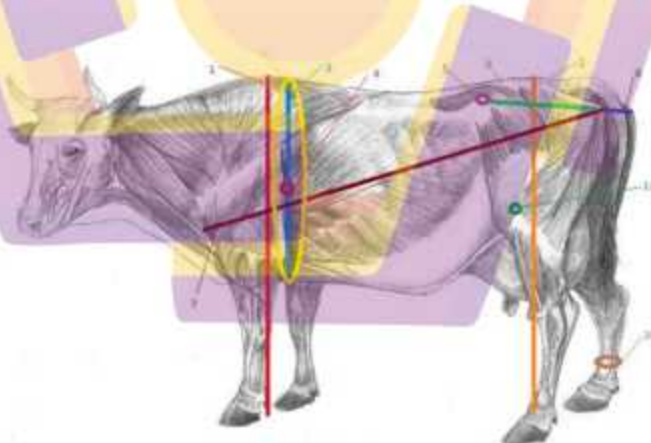
BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Dataset Penelitian

4.1.1 Preprocessing Data

Berdasarkan hasil pengumpulan data yang dilakukan dari Agustus 2008 hingga September 2020 terkumpul 1.523 sapi Hereford. Sapi Hereford milik dua peternakan swasta: (1) dengan pakan terkonsentrasi di wilayah Nizhny Novgorod, Rusia, dan (2) dengan pakan biji-bijian di wilayah Voronezh, Rusia (Ruchay, 2021). Setelah dilakukan berbagai tahapan *data labelling*, *data cleaning*, *data normalization*, dan *fitur selection*. Data yang digunakan dalam penelitian yaitu 1523 data yang memiliki tiga belas variabel.



Gambar 4.1. Dimensi Tubuh Sapi

Pada gambar 4.1 terlihat dimensi masing-masing sapi seperti : (1) withers height, (2) hip height, (3) chest depth, (4) chest width, (5) ilium width, (6) sciatic

tubercle width, (7) oblique body length, (8) oblique rear length, (9) heart girth, (10) metacarpus girth, (11) backside half-girth. Pengukuran tubuh dilakukan secara manual oleh seorang ahli dengan menggunakan pita pengukur biasa atau tongkat Lydlin atau kompas Wilkens dan dicatat dalam sentimeter (Ruchay, 2021).

Berbeda dengan penelitian yang dilakukan (Ruchay et al., 2022) pada sapi, menganalisis genetik dan fenotipik menggunakan uji heterozigositas, analisis komponen utama, dan studi asosiasi genom-wide (GWAS). Mereka juga menciptakan sistem tangkapan gambar 3-D untuk mengukur fitur eksterior sapi. Hasil penelitian menunjukkan korelasi antara fitur eksterior tertentu pada sapi dan produktivitas daging. Melalui analisis GWAS, penelitian ini mengidentifikasi wilayah signifikan dalam genom yang terkait dengan fitur-fitur tersebut. Temuan ini memberikan data berharga bagi komunitas penelitian dan mengusulkan metode baru untuk mengukur tubuh sapi tanpa kontak. Para peneliti juga mengidentifikasi varian genetik yang mendasari perbedaan dalam karakteristik kualitatif daging dengan menggunakan data terimputasi pada sekuens genom lengkap. Menemukan potensi lokus sifat kuantitatif pada kromosom sapi terkait dengan lebar dada, lingkaran dada, dan output daging pada tulang. Pembuatan database yang berisi gambar RGB-D mentah, awan titik, pengukuran berat hidup, data genotyping, dan karakteristik pemotongan untuk 96 sapi Aberdeen-Angus juga dilakukan. Para peneliti mengusulkan pendekatan inovatif untuk mengembangkan teknologi yang dapat menilai dan mengestimasi produktivitas hewan berdasarkan pengukuran tubuh otomatis menggunakan model 3-D sapi Aberdeen Angus dari beberapa kamera kedalaman dada dan output daging pada tulang sapi.

Penelitian ini (Gritsenko et al., 2023) mencakup pengembangan metode untuk mengestimasi produktivitas daging sapi jantan hitam-putih berdasarkan indeks penanda seperti berat hidup dan dimensi hewan saat lahir. Metode ini diuji di beberapa peternakan di wilayah Chelyabinsk dan menunjukkan tingkat akurasi berkisar antara 90 hingga 98%. Selain itu, studi ini juga membahas korelasi genetik dan perkiraan heritabilitas dari berbagai karakteristik terkait dengan produktivitas daging, dengan hasil menunjukkan bahwa metode dan algoritma yang diusulkan dapat menjadi solusi praktis untuk perusahaan pertanian.

Studi ini menemukan hubungan positif antara pengukuran tubuh hewan dan indeks berat hidup pada periode usia yang berbeda. Koefisien korelasi antara pengukuran saat lahir dan berat hidup pada usia 18 bulan berkisar antara 0,7 hingga 0,9, menunjukkan kemungkinan penggunaan pengukuran saat lahir untuk estimasi berat pra-potong. Penggunaan sistem tangkapan gambar RGB-D untuk mengukur berat hidup dan tinggi anak sapi dan induknya menunjukkan akurasi pengukuran yang tinggi. Selain itu, studi ini memanfaatkan sistem otomatis untuk mengukur parameter tubuh sapi hidup, berdasarkan rekonstruksi bentuk 3D non-rigid menggunakan data dari tiga kamera kedalaman. Sistem ini memiliki kesalahan pengukuran kurang dari 3%. Dengan penerapan model pembelajaran mesin, estimasi berat hidup sapi mencapai akurasi 95,67%.

Tabel 4.1. menunjukkan singkatan dan definisi dari hasil pengukuran tubuh sapi yang menjadi nilai variabel sebagai berikut:

Tabel 4.1. Singkatan dan Definisi Ukuran Tubuh Sapi

Singkatan	Definisi
<i>Withers Height (WH)</i>	Jarak vertikal dari titik tertinggi pada layu hingga titik tertinggi pada ujung jari kaki
<i>Hip Height (HH)</i>	Jarak vertikal dari titik tertinggi, tulang pinggul, ke titik terendah, tanah setinggi kaki belakang
<i>Chest Depth (CD)</i>	Jarak vertikal dari punggung ke pangkal ayah pada bagian yang paling jath jangkauan ayah
<i>Chest Width (CW)</i>	Lingkar badan pada suatu titik tepat di belakang kaki depan dan bahu serta tegak lurus terhadap sumbu badan
<i>Ilium Width (IW)</i>	Jarak antara titik terluar tulang ilium tegak lurus dengan pangkalnya
<i>Sciatic Tubercle Width (STW)</i>	Perbandingan dua titik sendi panggul yang tidak bergerak maju dengan cepat
<i>Oblique Body Length (OBL)</i>	Dari iskium posterior internal hingga ekstremitas humerus anterior
<i>Oblique Rear Length (OBR)</i>	Dari ujung posterior iskium interna hingga titik terluar ilium
<i>Heart Girth (HG)</i>	Bahu posterior tegak lurus dengan jarak punggung antara titik sudut
<i>Metacarpus Girth (MG)</i>	ketebalan pada titik tertipis tulang metakarpal
<i>Backside Half-Girth (BHG)</i>	jarak dari tonjolan lateral sendi lutut kiri ke belakang di bawah ekor dan ke titik yang sama pada sendi kanan

Proses Tabel 4.1. berupa kumpulan data berisi peta kedalaman gambar RGB yang disinkronkan dari tampilan kiri, kanan, dan atas (ukuran kedalaman dan gambar RGB masing-masing adalah 512 x 424 dan 1920 x 1080 piksel), matriks transformasi, pengukuran manual dan stempel waktu pada sapi (Ruchay et al., 2020). Berbeda dengan penelitian (Weber et al., 2020) proses pengumpulan data sapi dilakukan melalui pengambilan gambar 2D dari area punggung sapi Nellore. Gambar-gambar ini diambil menggunakan kamera yang dipasang di atas platform minum air sapi. Saat sapi sedang minum air, kamera akan mengambil gambar dari area punggung sapi.

(Weber et al., 2020) Data sapi yang digunakan dalam penelitian ini dikumpulkan dari 19 sapi jantan Nellore yang ditempatkan di dua feedlot yang berbeda di Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brasil. Gambar-gambar diambil antara pukul 06.00 pagi hingga 16.30 sore dari tanggal 8 Oktober hingga 20 November 2018.

(Diego et al., 2021) Data domba diperoleh melalui pengambilan gambar menggunakan *smartphone* Huawei P20 PRO dengan resolusi 1920 x 1080 piksel dan frame rate 60 frame per detik. Video-video tersebut direkam saat domba-domba tersebut ditimbang menggunakan timbangan elektronik. Dari setiap video, frame-frame diekstraksi dengan kecepatan 1 frame per detik, sehingga total terdapat 16 frame per video. Frame-frame ini kemudian diannotasi secara manual menggunakan perangkat lunak LabelMe untuk membuat masker yang memisahkan domba dari latar belakang.

Dataset akhir, yang diberi nama SHEEP32, terdiri dari 32 gambar yang diannotasi, masing-masing mewakili domba yang berbeda. Dataset ini mencakup informasi seperti identifikasi domba dan beratnya yang sesuai. Dataset yang telah diproses kemudian digunakan untuk ekstraksi fitur dan melatih model machine learning untuk mengestimasi berat domba (Diego et al., 2021).

4.2. Pemodelan Linear Regression dan Random Forest Regressor

Data yang sudah diolah yang terlihat pada Gambar 3.1 Cuplikan isi dataset tersebut kemudian dilakukan pemodelan Linear Regression dan Random Forest Regressor. Pada setiap proses *data labelling*, *data cleaning*, *data normalization*, *fitur selection* dan pengujian performa model *k-fold cross validation* kemudian di

hitung menggunakan evaluasi matrik nilai *Mean Absolute Error* (MAE), *Root Mean Absolute Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) dan *R-square*. Proses pemilihan fitur selection menggunakan 9 fitur variabel terbaik yaitu *age years*, *hip height*, *withers height*, *chest width*, *metacarpus girth*, *oblique rear length*, *oblique body length*, *ilium width*, *backside half-girth* dalam setiap model algoritma linear regression dan random forest regressor.

Hasil dari pemilihan fitur sembilan variabel berdasarkan hasil korelasi terbaik antar variabel yang digunakan dalam penelitian ini. Berbeda dengan penelitian (Weber et al., 2020) menggunakan metode *Correlation-based Feature Selection* (CFS). Metode ini digunakan untuk mengevaluasi dan memilih fitur yang memiliki hubungan kuat dengan variabel target (berat sapi) dan memiliki sedikit korelasi antar fitur. Metode ini menggabungkan evaluasi individual dari setiap atribut berdasarkan kemampuannya dalam mengestimasi berat sapi, serta tingkat redundansi antar atribut. Hasil dari pemilihan fitur ini adalah subset atribut yang paling relevan dan informatif untuk digunakan dalam analisis atau pemodelan (Weber et al., 2020).

Pada proses pengujian performa model menggunakan *k-fold cross validation* dengan repetisi acak untuk mendapatkan *split data* yang mampu memberikan nilai error yang terbaik dari setiap model algoritma linear regression dan random forest regressor yang digunakan. Berbeda dengan penelitian (Weber et al., 2020) menggunakan *nested 10-fold cross validation* dengan 100 repetisi acak. Repetisi acak dilakukan dengan mengacak ulang data sebelum dilakukan pembagian *fold*. Hal ini dilakukan untuk menghindari bias yang mungkin muncul akibat urutan data

yang tidak acak untuk memastikan hasil evaluasi yang lebih stabil dan akurat.

Proses cross-validation dilakukan (Diego et al., 2021) dengan menggunakan metode lima-fold cross-validation. Pada setiap iterasi, dataset dibagi menjadi lima bagian yang sama besar. Dari kelima bagian tersebut, empat bagian digunakan sebagai data pelatihan (80% dari dataset) dan satu bagian digunakan sebagai data pengujian (20% dari dataset). Hasil dari setiap iterasi kemudian diambil rata-ratanya untuk mendapatkan hasil akhir dari proses cross-validation.

Dalam penelitian (Jannah et al., 2023) dilakukan validasi silang dengan menggunakan 10-fold cross-validation. Hasil dari setiap iterasi cross-validation digunakan untuk menghitung error model regresi. Akurasi dihitung berdasarkan perbandingan antara berat badan yang diestimasi oleh model dengan berat badan sebenarnya pada data pengujian. Akurasi rata-rata dari 10 iterasi cross-validation digunakan sebagai estimasi akurasi model secara keseluruhan.

Root Mean Square Error (RMSE) yaitu dengan mengurangi nilai aktual (D) dengan nilai estimasi (P) kemudian dikuadratkan dan dijumlahkan (j) keseluruhan hasilnya kemudian dibagi dengan banyaknya data. Hasil perhitungan tersebut selanjutnya dihitung kembali untuk mencari nilai dari akar kuadrat, seperti dalam Rumus (4.3).

$$RMSE = \sqrt{\frac{\sum_{t=1}^j (D_t - P_t)^2}{j}} \quad (4.3)$$

Hasil estimasi pada P berdasarkan data aktual (D). Untuk menghitung nilai RMSE, maka nilai $(|D-P|^2)$ dibagi dengan (j), dimana (J) bernilai 100 dan diakarkan sehingga menghasilkan nilai akurasi RMSE 0.5 kg.

Berbeda dengan penyelesaian dari RMSE, pada Mean Absolute Error (MAE) ini cukup sederhana, yaitu dengan cara nilai estimasi (P) dikurangi nilai (D) yaitu $|D-P|$. Mean Absolute Error menunjukkan nilai kesalahan rata-rata dari nilai sebenarnya dengan nilai estimasi yaitu $|D-P|$, dihitung dalam Rumus (4.4).

$$MAE = \frac{1}{j} \sum_{t=1}^j |D_t - P_t| \quad (4.4)$$

Hasil peramalan di kolom P berdasarkan data aktual D. Untuk menghitung nilai MAE, maka nilai D-P dibagi dengan j, dimana j bernilai 100 dan diakarkan sehingga menghasilkan nilai MAE terbaik sebesar 0.5 kg.

Mean Absolute Percentage Error (MAPE) adalah dengan melakukan penjumlahan secara keseluruhan dengan terlebih dahulu melakukan pengurangan nilai data aktual (D) dengan data estimasi (P) kemudian membaginya dengan data aktual (D) diharuskan nilainya absolut ($|D-P|$) dan dikalikan dengan 100 kemudian dibagi dengan banyaknya data (j). Hasil perhitungan tersebut selanjutnya dihitung Kembali untuk mencari nilai dari akar kuadrat, dihitung dalam Rumus (4.5).

$$MAPE = \frac{\sum_{t=1}^j \left| \left(\frac{D_t - P_t}{D_t} \right) 100 \right|}{j} \quad (4.5)$$

Hasil peramalan di P berdasarkan data aktual (D). Untuk menghitung nilai MAPE, maka nilai error absolut dibagi dengan data actual (D) dan dikali 100 dan di bagi j, dimana j bernilai 100 sehingga menghasilkan nilai MAPE terbaik sebesar 0.07%.

Besar nilai dari koefisien determinasi adalah berkisar $0 \leq R^2 \leq 1$. Jika R^2 mendekati 1 maka dapat dikatakan pengaruh variabel bebas terhadap variabel terikat adalah besar yang berarti model yang digunakan baik untuk menjelaskan

Hasil korelasi menunjukkan adanya hubungan yang nyata antara bobot badan ternak (bobot hidup) dengan berbagai dimensi tubuh sapi yang diukur. Korelasi tertinggi terjadi antara bobot badan dengan Age Years sebesar 0.62 yang menunjukkan bahwa sapi dengan Age Years lebih besar cenderung memiliki bobot badan lebih tinggi. Data yang diperoleh menunjukkan korelasi yang menarik antara variabel-variabel tersebut. Penelitian ini mengeksplorasi korelasi beberapa dimensi pengukuran pada domba dengan menggunakan model regresi untuk menghasilkan estimasi yang sangat akurat. Pentingnya performa dan manajemen ternak untuk perhitungan bobot domba (Zaenab et al., 2023), pada penelitiannya menerapkan metode multiple linear regression. Dari hasil penelitian didapatkan nilai MAE terkecil sebesar 0,883 dan error sebesar 77,40%. Dengan melakukan kombinasi variabel berdasarkan nilai koefisien pada model persamaan regresi, body length (BL), chest girth (CG), and shoulder height (SH) dihasilkan nilai body weight (BW) = $67,004 + 0.589BL + 0.453CG + 0.399SH$. Namun, untuk nilai kombinasi yang lainnya masih bisa ditingkatkan.

4.3.2 Data Normalization

Gambar 4.3 menunjukkan *pseudocode* model *data normalization* proses untuk mengubah data menjadi bentuk standar, sehingga memudahkan dalam mengolah dan menganalisis dataset sapi. Normalisasi data bertujuan untuk memastikan bahwa setiap atribut mempunyai kontribusi yang seimbang untuk memperoleh hasil penelitian yang lebih akurat.


```
# Step 3: Normalize the data
scaler = StandardScaler()
X = scaler.fit_transform(X)
```

Gambar 4.3. Pseudocode Data Normalization

Terlihat pada Gambar 4.3. *Pseudocode Data Normalization* mengacu pada penggunaan fungsi `StandardScaler()` dari pustaka `scikit-learn` dalam bahasa pemrograman Python untuk melakukan normalisasi data. Dalam langkah pertama, objek `scaler` dibuat dengan menggunakan `StandardScaler()`. Selanjutnya, metode `fit_transform()` dari objek `scaler` digunakan untuk melakukan normalisasi data pada matriks `X`. Proses normalisasi ini melibatkan menghitung rata-rata dan standar deviasi dari setiap fitur dalam dataset, dan kemudian mengubah setiap nilai fitur ke dalam bentuk yang memiliki rata-rata nol dan standar deviasi satu. Dengan melakukan normalisasi ini, data diubah ke dalam skala yang seragam, memfasilitasi analisis statistik atau penggunaan model machine learning yang memerlukan data dalam skala yang seimbang dan terstandarisasi.

Data Sesudah Normalisasi:

	0	1	2	3	4	5	6	\
0	0.612004	1.051534	1.234804	0.446088	-1.032622	0.743810	0.124084	
1	-1.633977	-1.505999	-0.247756	-0.909770	-1.644858	-0.790487	-0.094474	
2	0.612004	-0.979244	-0.989076	0.296131	-1.032622	-0.108577	-0.021621	
3	-1.633977	-2.031555	-2.224609	-1.060508	-0.879563	0.960964	-0.094474	
4	-1.633977	0.336145	0.246458	0.748343	-1.338740	0.232378	-0.021621	
...	...	...	...	...	...	...	...	
1518	0.612004	0.073068	0.493564	0.597606	0.004088	-0.108577	-0.094474	
1519	0.612004	0.336145	-0.000649	-0.759033	0.497970	1.084765	-0.240179	
1520	0.612004	0.073068	-0.247756	-0.759033	2.020561	1.084765	-0.167327	
1521	0.612004	0.336145	-0.000649	-0.306820	2.793857	1.084765	-0.167327	
1522	0.612004	0.336145	0.246458	-0.608295	2.181621	0.232378	-0.167327	

Gambar 4.4. Cuplikan Data Normalization

Pada gambar 4.4 terlihat cuplikan hasil *data normalization* untuk memastikan bahwa fitur `X` dalam dataset memiliki rentang nilai yang serupa. Normalisasi juga

dapat membantu mengurangi dampak outlier dalam data, karena outlier mungkin memiliki dampak yang lebih besar pada model jika fitur-fitur memiliki skala yang berbeda-beda.

4.3.3 Feature Selection

Gambar 4.5 menunjukkan *pseudocode* model *performa feature selection* pola hubungan untuk mendapatkan nilai variabel yang terbaik.

```
# Step 4: Perform feature selection (if necessary)
Xnew = X
```

Gambar 4.5. *Pseudocode Feature Selection*

Terlihat dari Gambar 4.5. *feature selection* variabel X dengan 9 fitur variabel terbaik yaitu *age years*, *hip height*, *withers height*, *chest width*, *metacarpus girth*, *oblique rear length*, *oblique body length*, *ilium width*, *backside half-girth* untuk mendapatkan nilai error yang terbaik. Hasil dari proses *feature selection* yang dilakukan pada sebuah model dengan sembilan variabel tujuan utama dari penelitian ini adalah untuk memilih jumlah variabel yang menghasilkan tingkat error terkecil.

Proses ini sangat penting karena dapat mengurangi kompleksitas model, meningkatkan interpretabilitas, serta menghemat sumber daya komputasi. Hasil penelitian ini menunjukkan bahwa dengan menggunakan sembilan variabel yang paling relevan, tingkat error model tetap optimal. Dalam kata lain, pengurangan jumlah variabel tidak mengorbankan error estimasi, tetapi dapat mempermudah interpretasi model. Pentingnya pemilihan variabel yang tepat dalam menghasilkan

model yang efisien dan efektif. Ini adalah kontribusi yang berharga dalam pengembangan ilmu pengetahuan dan aplikasi praktis di berbagai bidang, dari ilmu data hingga pengambilan keputusan berbasis bukti.

Hasil penelitian ini menunjukkan bahwa tidak selalu lebih banyak variabel akan menghasilkan hasil yang lebih baik; sebaliknya, pemilihan variabel yang cerdas dan terfokus adalah kunci untuk meraih tingkat akurasi yang optimal. Temuan ini memberikan panduan berharga untuk menyederhanakan model dan meningkatkan keefektifan analisis data.

4.3.4 K-Fold Cross-Validation

Gambar 4.6. menunjukkan model *k-fold cross-validation* untuk mendapatkan pengujian performa model algoritma yang andal dan akurat.

```
# Initialize K-fold cross-validation
kf = KFold(n_splits=10, shuffle=True, random_state=42)
```

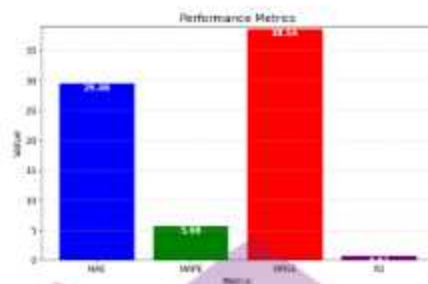
Gambar 4.6. *Pseudocode K-Fold Cross-Validation*

Terlihat dari Gambar 4.6. *k-fold cross-validation* splits data optimal sebanyak 10 kali lipat untuk mendapatkan nilai split yang terbaik. K-fold cross-validation digunakan dalam evaluasi kinerja model. Namun, menentukan nilai K yang optimal adalah langkah krusial dalam proses ini, dan penelitian ini berusaha untuk menjawab pertanyaan seputar berapa nilai K yang paling efektif. Hasil penelitian ini menunjukkan bahwa melalui serangkaian eksperimen yang cermat dan teliti, nilai K yang optimal adalah antara $n=10$, dengan melakukan split data sebanyak n kali lipat. Ini adalah temuan yang sangat menarik, karena mengindikasikan bahwa

membagi data menjadi 10 kelompok yang berbeda memberikan hasil yang paling optimal dalam evaluasi model. Pilihan ini bukan hanya berdampak pada error estimasi model, tetapi juga memberikan kerangka kerja yang efisien dalam memanfaatkan data secara maksimal untuk melatih dan menguji model. Dengan memilih nilai K yang optimal, penelitian ini mengilustrasikan bagaimana metode k-fold cross-validation dapat dioptimalkan untuk menghasilkan hasil yang paling andal dan efektif.

4.3.5 Algoritma Linear Regression

Setelah data 1523 sapi Hereford melalui *preprocessing* berikutnya adalah menggunakan metode *Linear Regression*. Ini menunjukkan hasil kinerja metrik dengan sembilan variabel menggunakan algoritma regresi linier pada Gambar 4.7. Metode LR ini dapat mempercepat data dalam hasil proses estimasi data. Data akan dimasukkan ke dalam google collabs untuk melakukan pengkodean data LR. Data tersebut akan mendapatkan hasil nilai *Mean Absolute Error* (MAE), *Root Mean Absolute Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) dan *R-square* yang sudah di dapat melalui proses pengkodean data LR. Berikut hasil pengkodean data LR pada Gambar 4.7. di bawah ini:



Gambar 4.7. Kinerja Metrik Algoritma Linear Regression dengan Sembilan Variabel dan 10-Fold Cross Validation

Dalam perjalanan eksplorasi melalui algoritma Linear Regression dengan mempertimbangkan sembilan variabel yang relevan temuan yang signifikan dan berarti terkuak di balik estimasi Gambar 4.7 hasil penelitian ini yang mengeksplorasi tingkat error model dengan cermat. Analisis terperinci menunjukkan bahwa model ini mampu menghasilkan hasil yang sangat dekat dengan data aktual, dengan nilai Mean Absolute Error (MAE) sebesar 29.460 kg. Angka ini menjadi penanda ketepatan dan kemampuan model dalam estimasi. Sementara itu, nilai Mean Absolute Percentage Error (MAPE) sebesar 5.678% menjadi refleksi keakuratan model dalam memperkirakan variabilitas. Selanjutnya, Root Mean Square Error (RMSE) yang tercatat hanya sebesar 38.564 kg memberikan gambaran menyeluruh tentang ketepatan estimasi model. Keberhasilan model dalam meminimalkan kesalahan estimasi semakin ditegaskan melalui nilai R-Square yang mencapai 0.622, menciptakan sebuah indikator yang sangat tinggi untuk kecocokan model. Hal ini membuktikan bahwa model Linear Regression dengan sembilan variabel mampu menjelaskan sebagian besar variabilitas dalam data dengan sangat baik.

Hasil ini menjadi bukti konkret atas kekuatan model dalam menganalisis hubungan antar variabel yang digunakan, membawa penelitian ini ke tingkat kesuksesan yang signifikan dalam menggambarkan dan memahami kompleksitas data dengan pendekatan yang matang.

Berbeda dengan penelitian oleh (Weber et al., 2020) mengembangkan model estimasi berat badan sapi Girolando menggunakan pengukuran tubuh dan citra yang diekstraksi dari citra sapi. Algoritma SVM memperoleh hasil terbaik menghasilkan nilai R^2 sebesar 0,70, RMSE sebesar 42,52 kg dan koefisien korelasi sebesar 0,71.

Penelitian yang dilakukan oleh (Ibrahim et al., 2021) untuk mengestimasi korelasi antara pengukuran tubuh (panjang tubuh, lingkar dada, dan tinggi pundak) dengan bobot pada domba Batur betina. Menghasilkan nilai besaran korelasi tertinggi body weight (BW) 0.866, body length (BL) 0.433 dan shoulder height (SH) 0.369 kombinasi estimasi terbaik yaitu panjang tubuh dan lingkar dada. Rumus regresi terbaik untuk mengestimasi berat badan adalah $BW = -56.522 + 0.509BL + 0.843CG$.

Gambar 4.8, menunjukkan model algoritma LR untuk mendapatkan nilai estimasi yang optimal.

```

for train_index, test_index in kf.split(X_new):
    X_train, X_test = X_new[train_index], X_new[test_index]
    y_train, y_test = y[train_index], y[test_index]

    # Fit the model
    model = LinearRegression()
    model.fit(X_train, y_train)

    # Predict
    y_pred = model.predict(X_test)

    # Calculate evaluation metrics
    mae = mean_absolute_error(y_test, y_pred)
    mse = mean_squared_error(y_test, y_pred)
    rmse = np.sqrt(mse)
    mape = np.mean(np.abs((y_test - y_pred) / y_test)) * 100
    corr, _ = pearsonr(y_test, y_pred)
    r2 = r2_score(y_test, y_pred)

```

Gambar 4.8. Pseudocode Model Algorithm Linear Regression

Gambar 4.8. menunjukkan dalam pengembangan model estimasi yang canggih, langkah-langkah ini membawa melalui perjalanan analisis yang menarik. Dengan menggunakan teknik validasi silang k-fold, dataset dibagi menjadi set pelatihan dan pengujian, memastikan keakuratan dan ketangguhan model Linear Regression. Melalui langkah-langkah ini, model di-fit pada data pelatihan dan kemudian diuji pada data pengujian, menghasilkan estimasi yang sangat diperhitungkan.

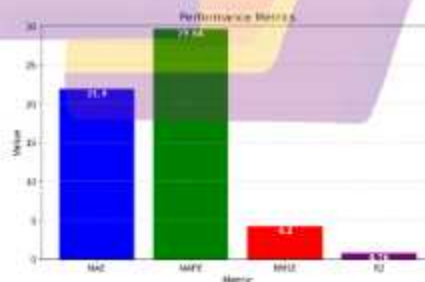
Model Linear Regression digunakan sebagai landasan utama, membuka pintu untuk pemahaman mendalam tentang hubungan antara variabel input dan output. Melalui estimasi yang dihasilkan, evaluasi model dilakukan dengan mempertimbangkan metrik seperti Mean Absolute Error (MAE), Mean Squared

Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), korelasi Pearson, dan R-squared. Hasil dari evaluasi ini memberikan pandangan yang holistik tentang sejauh mana model mampu mengestimasi output dengan akurat.

Hasil penelitian ini akan menjadi landasan penting dalam pengembangan model dan penelitian lanjutan di berbagai bidang, dari ilmu data hingga pembelajaran mesin. Penelitian ini memperkaya pemahaman tentang praktik terbaik dalam validasi model dan akan memandu pengambilan keputusan yang lebih cerdas dalam penggunaan data untuk analisis dan pengembangan model

4.3.6 Algoritma Random Forest Regressor

Berikutnya adalah menggunakan metode Random Forest Regressor. Ini menunjukkan hasil kinerja metrik dengan Sembilan variabel menggunakan algoritma random forest regressor pada Gambar 4.9. Data tersebut akan mendapatkan hasil MAE, RMSE, MAPE dan R-square yang sudah di dapat melalui proses pengkodean data RFR. Berikut hasil pengkodean data RFR pada Gambar 4.9 di bawah ini:



Gambar 4.9. Kinerja Metrik Algoritma Random Forest Regressor dengan Sembilan Variabel dan 10-Fold Cross Validation

Gambar 4.9. menunjukkan bahwa algoritma random forest regressor dengan sembilan variabel menunjukkan tingkat error nilai MAE sebesar 21.929 kg, model ini menunjukkan kemampuan untuk mengestimasi nilai yang mendekati dengan baik terhadap data aktual. Meskipun MAE ini mewakili tingkat kesalahan yang lebih tinggi dibandingkan dengan beberapa metode lainnya, tetapi perlu diingat bahwa setiap penelitian memiliki karakteristik unik dan penanganan variabilitas yang berbeda. Nilai MAPE sebesar 29.433% adalah indikator yang mengesankan dari tingkat error model ini. Ini menandakan bahwa model memiliki kemampuan yang sangat baik dalam memberikan estimasi yang sangat tepat, bahkan pada skala data yang besar. Meskipun nilai RMSE sebesar 4.407 kg menunjukkan adanya tingkat ketidakpastian dalam estimasi, perlu diingat bahwa setiap model memiliki batasan dan faktor-faktor yang mempengaruhinya. Dalam konteks ini, RMSE yang relatif rendah masih mencerminkan nilai error yang dapat diterima dalam estimasi.

Pentingnya temuan ini juga tercermin dalam nilai R-Square yang sangat tinggi, sebesar 0.761. Ini adalah bukti yang sangat kuat bahwa model Random Forest Regressor ini mampu menjelaskan sebagian besar variabilitas dalam data. Hal ini menandakan bahwa lima variabel yang dipertimbangkan dalam penelitian ini memiliki pengaruh yang signifikan terhadap hasil yang diamati. Penelitian ini memberikan pandangan mendalam tentang kompleksitas hubungan antar variabel-variabel dan memperlihatkan potensi besar dari algoritma Random Forest Regressor dalam menganalisis fenomena yang kompleks.

Berbeda dengan hasil penelitian (A N Ruchay., et al 2021), menunjukkan bahwa algoritma pembelajaran mesin, khususnya RandomForestRegressor,

memberikan hasil terbaik dalam estimasi berat badan sapi Hereford nilai MAE sebesar 24.965, RMSE sebesar 37.259, MAPE sebesar 4.853 dan R2 sebesar 0.644 berdasarkan berbagai ukuran morfologis dan usia. Namun, hasil nilai MAE dengan penerapan algoritma machine learning masih bisa ditingkatkan.

Gambar 4.10. menunjukkan model algoritma RFR untuk mendapatkan nilai estimasi yang akurat.

```
for train_index, test_index in kf.split(X_new):
    X_train, X_test = X_new[train_index], X_new[test_index]
    y_train, y_test = y[train_index], y[test_index]

    # Fit the model
    model = RandomForestRegressor(n_estimators=100, random_state=42)
    model.fit(X_train, y_train)

    # Predict
    y_pred = model.predict(X_test)

    # Calculate evaluation metrics
    mae = mean_absolute_error(y_test, y_pred)
    mse = mean_squared_error(y_test, y_pred)
    rmse = np.sqrt(mse)
    r2 = r2_score(y_test, y_pred)
```

Gambar 4.10. Pseudocode Model Algorithm Random Forest Regressor

Terlihat gambar 4.10. menunjukan dalam proses pelatihan model, kode di atas menggunakan pendekatan validasi silang K-Fold untuk mengukur kinerja model Random Forest Regressor. Setiap iterasi dari K-Fold membagi data menjadi set pelatihan dan pengujian, kemudian model diperbarui dan dievaluasi. Hasil estimasi kemudian diukur dengan metrik evaluasi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE), serta koefisien determinasi R-squared (R2). Pendekatan ini memungkinkan penilaian yang komprehensif terhadap performa model terhadap data pengujian, memberikan wawasan yang penting dalam validitas dan kehandalan model tersebut.

4.3.7 Nilai Hasil Error Model Algoritma

Hasil penelitian menunjukan banyak variasi model algoritma yang sudah dilakukan mulai menggunakan 1523 data sapi Hereford yang dapat dilihat secara keseluruhan pada Tabel 4.2. Berikut adalah nilai hasil error model algoritma:

Tabel 4.2. Nilai Hasil Error Model Algoritma

No.	Algoritma	Jumlah Dataset	Feature Selection Variabel	K-Fold	MAE	MAPE	RMSE	R2
1.	Linear Regression	1523	9	10-fold	29.460	5.678	38.564	0.622
2.	Random Forest Regressor				21.902	4.201	29.433	0.761

Tabel 4.2. menunjukkan hasil nilai *Feature Selection* dan pengujian performa menggunakan *10-Fold Cross Validation* memainkan peran penting dalam mengukur kinerja estimasi model pembelajaran mesin, terutama terlihat dalam konteks algoritma Random Forest Regressor. Hasil kontribusi penelitian yang cermat, yang melibatkan sejumlah variabel terpilih, mengungkapkan pola menonjol yang menggarisbawahi pentingnya proses ini. Peneliti menggunakan 9 variabel yang dipilih memiliki nilai hubungan korelasi yang tinggi dibandingkan menggunakan 12 fitur variabel, 9 fitur variabel terbaik yaitu *age years*, *hip height*, *withers height*, *chest width*, *metacarpus girth*, *oblique rear length*, *oblique body length*, *ilium width*, *backside half-girth* dengan cermat dan melakukan validasi silang 10 kali lipat pada model algoritma dapat memberikan nilai MAE mencapai

titik terkecil yang mengesankan yaitu 21.902 kg, ditambah dengan MAPE sebesar 4.201%, menyoroti ketepatan estimasi. Hasil ini diperkuat lebih lanjut dengan RMSE sebesar 29.433 kg dan nilai R-square yang sangat tinggi sebesar 0.761, yang menunjukkan model yang menangkap pola mendasar dengan akurasi yang optimal.

Karena penelitian ini menggunakan algoritma random forest regressor dengan jumlah data yang besar dan pemilihan fitur yang tepat dilihat dari nilai *Correlation Heatmap* dapat memberikan nilai error yang lebih baik dibandingkan dengan penelitian yang dilakukan (A N Ruchay., et al 2022), dengan menggunakan algoritma pembelajaran mesin terbaru, khususnya ExtraTreesRegressor, penelitian ini berhasil mengungkap bahwa metode ini mampu memberikan estimasi berat badan sapi Hereford yang paling akurat. Hasilnya, diukur dengan menggunakan beberapa metrik evaluasi kinerja, menunjukkan tingkat akurasi yang luar biasa, dengan Mean Absolute Error (MAE) sebesar 22.929, Root Mean Squared Error (RMSE) sebesar 29.641, Mean Absolute Percentage Error (MAPE) sebesar 4.407, dan koefisien determinasi (R^2) sebesar 0.758. Penelitian ini dilakukan dengan mempertimbangkan berbagai ukuran morfologis dan usia sapi, menggambarkan kesempurnaan algoritma ini dalam menangani keragaman data dari berbagai aspek.

Kemudian algoritma Linear Regression memberikan hasil nilai MAE sebesar 29.460, nilai MAPE sebesar 5.678, nilai RMSE 38.564, dan R^2 sebesar 0.622. Karena keterbatasan dalam menangani hubungan nonlinier meskipun regresi linear dapat dilengkapi dengan transformasi nonlinier dari variabel independen, dalam penelitian ini Linear Regression hanya dapat menangkap pola nonlinier yang sederhana dan tidak dapat menangani pola nonlinier yang lebih kompleks yaitu

dengan 9 fitur variabel dan 1523 data yang digunakan dalam penelitian ini. Sehingga pola nonlinier yang kompleks dan jumlah dataset yang besar dapat mempengaruhi hasil algoritma linear regression.

Berbeda dengan penelitian yang dilakukan (Weber et al., 2020) bahwa algoritma Bagging memberikan hasil terbaik dengan MAE sebesar 13,44 kg ($\pm 2,76$), RMSE sebesar 15,88 kg ($\pm 2,86$), MAPE sebesar 2,27%, dan koefisien korelasi sebesar 0,75. Kemudian penelitian yang dilakukan oleh (Diego et al., 2021) menunjukkan bahwa model yang dikembangkan memiliki Mean Absolute Error (MAE) sebesar 3.099 kg dan Adjusted R^2 sebesar 0.687.

Penelitian (Huma et al., 2019) menunjukkan bahwa metode random forests menghasilkan nilai MAE sebesar 1.242, RMSE sebesar 2.129, koefisien korelasi sebesar (0.947-0.994) dan koefisien determinasi yang tinggi (0.896-0.988). Selanjutnya experiment (Ruchay et al., 2022) menggunakan algoritma stacking regressor memberikan performa terbaik dalam mengestimasi bobot babi menghasilkan MAE sebesar 4.331 dan MAPE sebesar 4.296 estimasi bobot sapi yang akurat memberikan keunggulan dalam manajemen stok, memungkinkan peternak membuat keputusan tepat terkait waktu panen, pemilihan ternak, dan strategi manajemen lainnya.

BAB V

PENUTUP

5.1. Kesimpulan

1. Hasil penelitian menunjukkan bahwa model random forest regressor mampu mencapai nilai error yang lebih kecil, dengan MAE sebesar 21.902 kg, MAPE sebesar 4.201%, RMSE sebesar 29.433 kg, dan R-square sebesar 0.761. Sebaliknya, model linear regression menunjukkan nilai error MAE sebesar 29.460, nilai MAPE sebesar 5.678, nilai RMSE 38.564, dan R2 sebesar 0.622. Hasil ini memberikan wawasan yang signifikan bahwa, dalam konteks penelitian ini, model random forest regressor lebih unggul dalam memberikan estimasi yang lebih akurat dibandingkan dengan model linear regression.
2. Dalam penelitian ini, penggunaan model algoritma random forest regressor berhasil menghasilkan nilai Mean Absolute Error (MAE) yang cukup rendah, tepatnya sebesar 21.902 kg. Hasil ini menunjukkan bahwa model mampu memberikan estimasi dengan tingkat error yang kecil dalam memperkirakan nilai variabel. Keberhasilan mencapai MAE sebesar 21.902 kg mengindikasikan bahwa algoritma random forest regressor dapat secara efektif menangkap pola dan hubungan yang mendasari data, memberikan kontribusi yang signifikan dalam konteks prediktif.
3. Penelitian ini menghasilkan temuan yang menarik dengan menunjukkan bahwa terdapat korelasi tertinggi sebesar 0.62 antara bobot badan dan

umur sapi. Temuan ini mengindikasikan bahwa sapi yang memiliki umur sapi lebih besar cenderung memiliki bobot badan yang lebih tinggi. Hasil korelasi yang kuat ini memberikan wawasan yang signifikan terkait hubungan antara dua variabel tersebut dalam konteks penelitian ini. Dengan demikian, pemahaman tentang korelasi ini dapat memiliki implikasi penting dalam pemilihan dan peningkatan kualitas sapi untuk tujuan pengembangan peternakan atau manajemen ternak secara efektif.

4. Penelitian ini menunjukkan bahwa pemilihan fitur dan pengujian performa menggunakan *k-fold cross validation* memiliki peran yang signifikan dalam meningkatkan kinerja prediktif model pembelajaran mesin, khususnya dalam konteks algoritma Random Forest Regressor. Dengan melakukan penelitian yang cermat dan memperhatikan sejumlah variabel terpilih dan jumlah *k-fold cross validation*, hasil menunjukkan adanya pola menonjol yang menekankan pentingnya proses penelitian yang dilakukan.

5.2. Saran

Ada beberapa hal yang penulis rekomendasikan untuk dilakukan peneliti lain pada topik penelitian yang sama, antara lain:

1. Menambah jumlah dataset jenis sapi lokal yang berbeda untuk melihat nilai error model yang dihasilkan.
2. Pada pre-processing masih mengolah data, peneliti selanjutnya dapat mengolah data berupa image pre-processing 2D dan 3D.
3. Melakukan peningkatan model yang digunakan seperti *feature selection* dan *k-fold cross validation*.

4. Menambah model algoritma seperti *support vector regression*, *k-neighbors regressor*, *multi-layer perceptron regressor*, *gradient boosting regression*, *light gradient boosting machine*, dan *extreme gradient boost regressor* untuk melihat dampaknya error model terbaik yang digunakan.
5. Melakukan penelitian estimasi bobot sapi untuk kombinasi hubungan antar variabel yang memiliki nilai error model yang terbaik.
6. Pendekatan proses image pre-processing untuk mengklasifikasi kualitas daging sapi yang layak dikonsumsi.
7. Mengolah objek dataset gambar dan video untuk meningkatkan proses augmentasi estimasi bobot sapi secara *real time*.
8. Melakukan pengolahan dataset sapi lokal yang berbeda untuk mengestimasi usia sapi.
9. Melakukan *forecasting* untuk menentukan kategori harga jual sapi berdasarkan kualitas daging sapi.
10. Peneliti selanjutnya juga dapat melakukan penelitian untuk membuat sistem pakar mengenai diagnosa penyakit infeksius sapi menggunakan teknik computer vision dan machine learning.
11. Jika peneliti selanjutnya menghasilkan nilai akurasi model yang lebih baik dapat menghasilkan aplikasi mobile yang bisa diakses oleh pengguna.

DAFTAR PUSTAKA

PUSTAKA BUKU

- Groß, J. (2003). Linear regression (Vol. 175). Springer Science & Business Media.
- Weisberg, S. (2005). Applied linear regression (Vol. 528). John Wiley & Sons.
- Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: Random forest. In Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings 3 (pp. 246-252). Springer Berlin Heidelberg.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons.

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- Ashari, Latif, N., & Astuti, A. D. (2019). Pengolahan citra digital untuk menentukan bobot sapi menggunakan metode Canny edge detection. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, 5(1), 1–6. <https://doi.org/10.35329/jiik.v5i1.24>
- Ashwini, J. P. (2019). Prediction of Body Weight based on Body Measurements in Crossbred Cattle. *International Journal of Current Microbiology and Applied Sciences*. <https://doi.org/10.20546/ijcmas.2019.803.186>
- Azis, R., Ciptadi, G., Wahjuningsih, S., Hariyono, D. N. H., Tribudi, Y. A., & Nurgiartiningsih, V. M. A. (2023). Prediction of Body Weight from Body Measurements in Bali Cattle of Indonesia Using Regression Analysis. *Advances in Animal and Veterinary Sciences*, 11(9). <https://doi.org/10.17582/journal.aavs/2023/11.9.1486.1491>
- Beef From Farm To Table | Food Safety and Inspection Service. (n.d.). <https://www.fsis.usda.gov/food-safety/safe-food-handling-and-preparation/meat/beef-farm-table>
- Bezen, R., Edan, Y., & Halachmi, I. (2020). Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms. *Computers and Electronics in Agriculture*, 172, 105345. <https://doi.org/10.1016/j.compag.2020.105345>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>

- B. Sartono and U. D. Syafitri, "METODE POHON GABUNGAN: SOLUSI PILIHAN UNTUK MENGATASI KELEMAHAN POHON REGRESI DAN KLASIFIKASI TUNGGAL," *Forum Statistika Dan Komputasi*, vol. 15, no. 1, Jan. 2010, [Online]. Available: <https://journal.ipb.ac.id/index.php/statistika/article/download/4895/3328>
- Cabaneros, S. M., & Hughes, B. R. (2022). Methods used for handling and quantifying model uncertainty of artificial neural network models for air pollution forecasting. *Environmental Modelling and Software*, 158, 105529. <https://doi.org/10.1016/j.envsoft.2022.105529>
- Cang, Y., He, H., & Qiao, Y. (2019). An intelligent pig weights estimate method based on deep learning in sow stall environments. *IEEE Access*, 7, 164867–164875. <https://doi.org/10.1109/access.2019.2953099>
- Cominotte, A., Fernandes, A. F. A., Dórea, J. R. R., Rosa, G. J. M., Ladeira, M. M., Van Cleef, E. H. C. B., Pereira, G. L., Baldassini, W. A., & Neto, O. R. M. (2020). Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases. *Livestock Science*, 232, 103904. <https://doi.org/10.1016/j.livsci.2019.103904>
- Dakhlan, A., Saputra, A., Hamdani, M. D. I., & Sulastris, S. (2020). Regression Models and Correlation Analysis for Predicting Body Weight of Female Ettawa Grade Goat using its Body Measurements. *Advances in Animal and Veterinary Sciences*, 8(11). <https://doi.org/10.17582/journal.aavs/2020/8.11.1142.1146>
- Dallago, G. M., Wade, K. M., Cue, R., McClure, J. T., Lacroix, R., Pellerin, D., & Vasseur, E. (2021). Keeping Dairy Cows for longer: A Critical Literature Review on Dairy cow Longevity in High Milk-Producing Countries. *Animals*, 11(3), 808. <https://doi.org/10.3390/ani11030808>
- Dohmen, R., Çatal, Ç., & Liu, Q. (2021a). Computer vision-based weight estimation of livestock: a systematic literature review. *New Zealand Journal of Agricultural Research*, 65(2–3), 227–247. <https://doi.org/10.1080/00288233.2021.1876107>
- Dohmen, R., Çatal, Ç., & Liu, Q. (2021b). Image-based body mass prediction of heifers using deep neural networks. *Biosystems Engineering*, 204, 283–293. <https://doi.org/10.1016/j.biosystemseng.2021.02.001>
- Ghotbaldini, H., Mohammadabadi, M., Nezamabadi-pour, H., Babenko, O., Bushtruk, M., & Tkachenko, S. (2019). Predicting breeding value of body weight at 6-month age using Artificial Neural Networks in Kermani sheep breed. *Acta Scientiarum. Animal Sciences*, 41(1), 45282. <https://doi.org/10.4025/actascianimsci.v41i1.45282>

- Gupta, A. K., Singh, V., Mathur, P., & González, C. J. L. (2020). Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario. *Journal of Interdisciplinary Mathematics*, 24(1), 89–108. <https://doi.org/10.1080/09720502.2020.1833458>
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning. In Springer series in statistics. <https://doi.org/10.1007/978-0-387-84858-7>
- Home. (n.d.). National Cattlemen's Beef Association. <https://www.ncba.org/>
- Huma, Z. E., & Iqbal, F. (2019). Predicting the body weight of Balochi sheep using a machine learning approach. *Turkish Journal of Veterinary & Animal Sciences*, 43(4), 500–506. <https://doi.org/10.3906/vet-1812-23>
- Ibrahim, A., Artama, W. T., Budisatria, I. G. S., Yuniawan, R., Atmoko, B. A., & Widayanti, R. (2021). Regression model analysis for prediction of body weight from body measurements in female Batur sheep of Banjarnegara District, Indonesia. *Biodiversitas*, 22(7). <https://doi.org/10.13057/biodiv/d220721>
- Jannah, Z. N., Atmoko, B. A., Ibrahim, A., Harahap, M. A., & Panjono, P. (2023). Body weight prediction model analysis based on the body size of female Sakub sheep in Brebes District, Indonesia. *Biodiversitas*, 24(7). <https://doi.org/10.13057/biodiv/d240702>
- Y. A. Jatmiko, S. Padmadisastra, and A. Chadidjah, "ANALISIS PERBANDINGAN KINERJA CART KONVENSIONAL, BAGGING DAN RANDOM FOREST PADA KLASIFIKASI OBJEK: HASIL DARI DUA SIMULASI," *Media Statistika*, vol. 12, no. 1, p. 1, Jul. 2019, doi: 10.14710/medstat.12.1.1-12.
- Khorshidi-Jalali, M., Mohammadabadi, M., Esmailzadeh, A., Barazandeh, A., & Babenko, O. (2019). Comparison of artificial neural network and regression models for prediction of body weight in Raini Cashmere Goat. *Iranian Journal of Applied Animal Science*, 9(3), 453–461. <http://rep.btsau.edu.ua/bitstream/BNAU/3491/1/Comparison-%20of-Art.pdf>
- L. Breiman, "Classification and regression trees.," *Biometrics*, vol. 40, no. 3, p. 874, Sep. 1984, doi: 10.2307/2530946.
- Marai, I., El-Darawany, A. A., Fadiel, A., & Abdel-Hafez, M. (2007). Physiological traits as affected by heat stress in sheep—A review. *Small Ruminant Research*, 71(1–3), 1–12. <https://doi.org/10.1016/j.smallrumres.2006.10.003>
- Miller, G. A., Hyslop, J. J., Barclay, D., Edwards, A. J., Thomson, W., & Duthie, C. (2019). Using 3D imaging and machine learning to predict liveweight and

- carcass characteristics of live finishing beef cattle. *Frontiers in Sustainable Food Systems*, 3. <https://doi.org/10.3389/fsufs.2019.00030>
- Montgomery, J. M., Hollenbach, F. M., & Ward, M. D. (2012). Improving Predictions using Ensemble Bayesian Model Averaging. *Political Analysis*, 20(3), 271–291. <https://doi.org/10.1093/pan/mps002>
- Nordhausen, K. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second edition by Trevor Hastie, Robert Tibshirani, Jerome Friedman. *International Statistical Review*, 77(3), 482. https://doi.org/10.1111/j.1751-5823.2009.00095_18.x
- Paengkoum, S., Tatsapong, P., Taethaisong, N., Thongpea, S., Purba, R. a. P., & Paengkoum, P. (2021). Empirical evaluation and prediction of protein requirements for maintenance and growth of 18–24 months old Thai Swamp buffaloes. *Animals*, 11(5), 1405. <https://doi.org/10.3390/ani11051405>
- Rather, M. A., Bashir, I., Hamdani, A., Khan, N. N., Ahangar, S. A., & Nazki, M. (2020). Prediction of body weight from linear body measurements in Kashmir merino sheep. *Advances in Animal and Veterinary Sciences*, 9(2). <https://doi.org/10.17582/journal.aavs/2021/9.2.189.193>
- Rizaty, M. A. (2023, June 26). *Kebutuhan Hewan Kurban Capai 1,7 Juta Ekor saat Iduladha 2023*. *Dataindonesia.id*. <https://dataindonesia.id/agribisnis-kebutuhan/detail/kebutuhan-hewan-kurban-capai-17-juta-ekor-saat-iduladha-2023>
- Ruchay, A., Kober, V., Dorofeev, K., Колпаков, B. И., Gladkov, A., & Guo, H. (2022). Live weight prediction of cattle based on deep regression of RGB-D images. *Agriculture*, 12(11), 1794. <https://doi.org/10.3390/agriculture12111794>
- Ruchay, A., Kober, V., Dorofeev, K., Колпаков, B. И., & Мирошников, С. А. (2020). Accurate body measurement of live cattle using three depth cameras and non-rigid 3-D shape recovery. *Computers and Electronics in Agriculture*, 179, 105821. <https://doi.org/10.1016/j.compag.2020.105821>
- Ruchay, A., Гриценко, С. А., Ermolova, E., Bochkarev, A., Ermolov, S., Guo, H., & Pezzuolo, A. (2022). A comparative study of machine learning methods for predicting live weight of Duroc, Landrace, and Yorkshire pigs. *Animals*, 12(9), 1152. <https://doi.org/10.3390/ani12091152>
- Sant'Ana, D. A., Pache, M. C. B., Martins, J. a. L., Soares, W. P., De Melo, S. L. N., Garcia, V., Weber, V., Da Silva Heimbach, N., Mateus, R. G., & Pistori, H. (2021). Weighing live sheep using computer vision techniques and regression machine learning. *Machine Learning With Applications*, 5, 100076. <https://doi.org/10.1016/j.mlwa.2021.100076>

- Sherwin, G., Hyde, R., Green, M. A., Remnant, J., Payne, E., & Down, P. (2021). Accuracy of heart girth tapes in the estimation of weights of pre-weaned calves. *Vet Record Open*, 8(1). <https://doi.org/10.1002/vro2.16>
- Song, X., Bokkers, E., Van Mourik, S., Koerkamp, P. G., & Van Der Tol, P. (2019). Automated body condition scoring of dairy cows using 3-dimensional feature extraction from multiple body regions. *Journal of Dairy Science*, 102(5), 4294–4308. <https://doi.org/10.3168/jds.2018-15238>
- Sun, M., Hossain, M., Islam, T., Rahman, M., Hossain, M., & Hashem, M. (2020). Different body measurement and body weight prediction of jamuna basin sheep in Bangladesh. *SAARC Journal of Agriculture*, 18(1), 183–196. <https://doi.org/10.3329/sja.v18i1.48392>
- The State of Food and Agriculture 2010–11 Women in agriculture. Closing the gender gap for development. FAO Published in 2011, 247 pp. ISBN: 978-92-5-106768-0. Available at <http://www.fao.org/docrep/013/i2050e/i2050e00.htm> (also in Arabic, Chinese, French, Russian and Spanish). (2011). *Animal Genetic Resources Information*, 49, 118. <https://doi.org/10.1017/s2078633611000567>
- Weber, V. a. M., De Lima Weber, F., Da Silva Oliveira, A., Astolfi, G., Menezes, G. V., De Andrade Porto, J. V., Rezende, F. P. C., De Moraes, P. H., Matsubara, E. T., Mateus, R. G., De Araújo, T. L. a. C., Da Silva, L. O. C., De Queiroz, E. Q. A., De Abreu, U. G. P., Da Costa Gomes, R., & Pistori, H. (2020). Cattle weight estimation using active contour models and regression trees Bagging. *Computers and Electronics in Agriculture*, 179, 105804. <https://doi.org/10.1016/j.compag.2020.105804>
- Weleszczuk, J., Kosińska-Selbi, B., & Cholewińska, P. (2022). Prediction of Polish Holstein's economical index and calving interval using machine learning. *Livestock Science*, 264, 105039. <https://doi.org/10.1016/j.livsci.2022.105039>
- Worku, A. (2019). Body weight had highest correlation coefficient with heart girth around the chest under the same farmers feeding conditions for Arsi Bale sheep. *International Journal of Agricultural Science and Food Technology*, 006–012. <https://doi.org/10.17352/2455-815x.000035>
- Siami-Namini, S., & Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. *arXiv* (Cornell University). <https://doi.org/10.48550/arxiv.1803.06386>
- S. Raschka, Python Machine Learning. 2015. [Online]. Available: <https://www.amazon.com/Python-Machine-Learning-Sebastian-Raschka/dp/1783555130>

- Weber, V., De Lima Weber, F., Da Costa Gomes, R., Da Silva Oliveira, A., Menezes, G. V., De Abreu, U. G. P., De Souza Belete, N. A., & Pistori, H. (2020). Prediction of Girolando cattle weight by means of body measurements extracted from images. *Revista Brasileira De Zootecnia*, 49. <https://doi.org/10.37496/rbz4920190110>
- Roche, J., Friggens, N., Kay, J., Fisher, M. W., Stafford, K. J., & Berry, D. (2009). Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. *Journal of Dairy Science*, 92(12), 5769–5801. <https://doi.org/10.3168/jds.2009-2431>
- Andrew, W., Gao, J., Mullan, S., Campbell, N., Dowsey, A. W., & Burghardt, T. (2021). Visual identification of individual Holstein-Friesian cattle via deep metric learning. *Computers and Electronics in Agriculture*, 185, 106133. <https://doi.org/10.1016/j.compag.2021.106133>
- Hierons, R. M. (1999). Machine learning. Tom M. Mitchell. Published by McGraw-Hill, Maidenhead, U.K., International Student Edition, 1997, ISBN: 0-07-115467-1, 414 pages. Price: U.K. £22.99, soft cover. Software Testing, Verification & Reliability, 9(3), 191–193. [https://doi.org/10.1002/\(sici\)1099-1689\(199909\)9:3](https://doi.org/10.1002/(sici)1099-1689(199909)9:3)
- Kumar, S., & Singh, S. K. (2016). Automatic identification of cattle using muzzle point pattern: a hybrid feature extraction and classification paradigm. *Multimedia Tools and Applications*, 76(24), 26551–26580. <https://doi.org/10.1007/s11042-016-4181-9>
- Kumar, S., Singh, S. K., Singh, R. S., Singh, A. K., & Tiwari, S. (2016). Real-time recognition of cattle using animal biometrics. *Journal of Real-time Image Processing*, 13(3), 505–526. <https://doi.org/10.1007/s11554-016-0645-4>
- Kimura, A., Itaya, K., & Watanabe, T. (2004). Structural pattern recognition of biological textures with growing deformations: A case of cattle's muzzle patterns. *Electronics and Communications in Japan*, 87(5), 54–66. <https://doi.org/10.1002/ecjb.20076>
- Tharwat, A., Gaber, T., Hassanien, A. E., Hassanien, H. A., & Tolba, M. F. (2014). Cattle identification using muzzle print images based on texture features approach. In *Advances in intelligent systems and computing* (pp. 217–227). https://doi.org/10.1007/978-3-319-08156-4_22
- LeCun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Potter, D. C., & Gujarati, D. (2009). Basic Econometrics 5th Edition. McGraw-Hill. http://perpus.ekuitas.ac.id/index.php?p=show_detail&id=93996

- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001A>.
- Talukder and B. Ahammed, "Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh," *Nutrition*, vol. 78, p. 110861, Oct. 2020, doi: 10.1016/j.nut.2020.110861.
- Z. Mašetić and A. Subaşı, "Congestive heart failure detection using random forest classifier," *Computer Methods and Programs in Biomedicine*, vol. 130, pp. 54–64, Jul. 2016, doi: 10.1016/j.cmpb.2016.03.020.
- V. T. N. Chau and N. K. P. Phung, "Imbalanced educational data classification: An effective approach with resampling and random forest," *Proc. - 2013 RIVF Int. Conf. Comput. Commun. Technol. Res. Innov. Vis. Futur*, Nov. 2013, doi: 10.1109/rivf.2013.6719882.
- Ruchay, A., Kober, V., Dorofeev, K., Колпаков, В. И., Джуламанов, К. М., Kalschikov, V., & Guo, H. (2022). Comparative analysis of machine learning algorithms for predicting live weight of Hereford cows. *Computers and Electronics in Agriculture*, 195, 106837. <https://doi.org/10.1016/j.compag.2022.106837>.
- Ruchay, A., Колпаков, В. И., Kalschikov, V., Джуламанов, К. М., & Dorofeev, K. (2021). Predicting the body weight of Hereford cows using machine learning. *IOP Conference Series. Earth and Environmental Science*, 624(1), 012056. <https://doi.org/10.1088/1755-1315/624/1/012056>.

LAMPIRAN 1. DATASET SAPI

No.	Body Weight	Age Years	Withers Height	Hip Height	Chest Depth	Chest Width	Ilium Width	Sciatic Tubercle Width	Oblique Body Length	Oblique Rear Length	Heart Girth	Metacarpus Girth	Backside Half-Girth
1.	516	5	133	135	69	40	56	28	178	55	189	20	71
2.	358	3	121	129	60	36	47	25	153	47	173	18	65
3.	460	5	123	126	68	40	51	26	168	50	185	19	68
4.	314	3	119	121	59	41	46	25	142	44	171	19	66
5.	492	3	128	131	71	38	53	26	167	52	183	20	70
6.	455	5	122	124	65	48	51	26	161	50	188	20	69
7.	507	5	129	140	73	40	57	28	170	56	198	21	70
8.	427	5	123	125	68	39	50	27	156	51	193	19	69
9.	516	5	125	129	71	45	56	29	171	56	194	21	71
10.	422	5	122	119	68	40	52	27	147	53	185	19	67
11.	350	3	117	122	61	38	42	26	149	44	177	19	66
12.	511	5	120	126	69	45	51	28	157	53	193	20	70
13.	549	5	127	131	72	42	56	28	169	56	191	21	71
14.	398	3	117	122	68	36	47	25	160	48	179	19	68
15.	349	3	124	128	64	43	47	25	148	48	176	19	68
16.	414	3	117	118	60	41	51	26	155	47	177	20	69
17.	518	5	127	129	69	47	55	28	164	53	222	20	73
18.	478	5	124	129	71	42	53	26	166	54	205	19	68

19.	389	3	124	124	55	43	48	25	155	50	185	19	67
20.	316	3	120	125	61	36	45	25	145	46	167	18	67
21.	515	5	129	131	71	43	55	27	161	52	192	21	71
22.	445	3	122	128	67	49	50	26	152	48	191	20	68
23.	403	3	117	118	65	48	50	25	154	50	186	20	67
24.	460	5	127	129	72	41	58	26	161	55	189	19	67
25.	505	3	124	127	70	45	52	27	156	52	192	20	69
26.	407	3	120	127	65	39	51	26	150	49	178	19	68
27.	506	3	130	135	69	49	54	28	163	53	207	20	70
28.	488	3	126	127	68	54	52	27	166	51	184	19	69
29.	525	3	127	129	69	48	51	28	155	53	204	20	71
30.	710	5	135	137	75	50	66	29	189	60	231	22	75
31.	490	5	124	128	71	42	48	26	162	52	195	19	69
32.	473	3	126	127	67	45	50	27	159	49	183	19	70
33.	402	5	118	123	65	39	51	26	151	50	177	19	66
34.	458	5	122	123	68	41	53	27	158	50	186	20	68
35.	589	5	126	127	68	49	55	28	166	52	213	21	72
36.	481	5	123	131	68	45	52	27	159	51	190	20	69
37.	513	5	128	130	70	44	53	28	172	55	202	21	73
38.	365	3	120	126	61	42	45	26	152	47	174	19	65

39.	643	5	132	137	75	47	59	29	174	56	209	21	75
40.	584	5	134	139	73	44	56	28	183	55	202	20	73
41.	513	5	122	129	70	44	56	27	161	51	190	20	72
42.	589	5	132	133	72	44	45	28	171	52	194	20	74
43.	522	5	127	132	70	42	56	28	169	55	195	19	73
44.	536	5	128	130	71	40	58	29	172	54	196	20	72
45.	505	5	129	126	72	38	55	27	165	52	192	19	70
46.	447	3	122	130	68	37	50	26	158	49	183	19	68
47.	560	5	126	133	71	44	57	28	184	54	195	20	71
48.	467	3	125	127	70	45	51	26	164	54	204	19	69
49.	537	3	130	134	70	47	53	28	160	54	230	2	72
50.	493	5	126	127	67	46	49	26	166	52	190	19	68
51.	656	5	132	134	72	48	60	29	183	60	212	21	75
52.	412	3	122	123	68	37	51	27	142	52	184	18	68
53.	510	5	127	131	72	42	55	28	173	54	201	20	73
54.	437	3	118	120	70	38	50	26	160	46	191	19	67
55.	484	5	124	129	71	40	55	26	176	54	192	19	68
56.	629	5	131	139	70	43	56	29	180	55	203	21	74
57.	468	3	122	122	67	46	51	26	165	53	191	19	70
58.	440	3	120	121	66	40	50	27	160	51	181	19	69

59.	448	5	124	131	68	37	53	27	161	54	181	19	68
60.	378	3	125	130	66	35	50	25	164	50	173	19	67
61.	598	5	127	132	70	44	59	28	180	55	201	20	72
62.	502	3	122	127	72	40	52	27	167	50	188	19	70
63.	377	5	120	121	62	37	47	25	160	48	176	19	68
64.	455	5	124	131	70	40	53	26	173	52	193	20	70
65.	573	3	129	134	70	47	58	27	178	55	199	20	72
66.	362	3	117	122	60	37	48	26	149	47	177	19	68
67.	472	3	125	128	67	43	56	27	160	53	189	19	70
68.	445	5	131	132	69	41	51	27	172	52	188	19	69
69.	398	3	120	125	65	38	51	26	147	49	177	19	68
70.	479	5	119	125	63	41	51	26	165	53	184	19	68
71.	414	3	122	124	65	39	50	26	158	51	181	19	67
72.	446	3	125	129	71	38	53	27	161	55	195	19	68
73.	556	5	130	135	71	42	56	29	177	54	192	21	72
74.	371	3	117	120	62	37	50	26	150	47	178	19	66
75.	460	3	122	125	70	39	51	27	162	51	188	19	70
76.	493	3	124	128	69	41	52	27	160	52	193	20	70
77.	647	5	137	139	80	44	59	29	192	56	207	21	75
78.	432	3	118	120	69	40	52	26	153	53	188	19	68

79.	406	3	120	124	66	37	49	26	157	49	178	19	68
80.	427	3	120	120	66	37	50	26	163	49	192	20	68
81.	470	3	127	131	67	40	54	26	158	53	186	20	69
82.	371	3	120	126	65	41	49	25	147	48	181	19	65
83.	474	5	121	126	69	45	47	26	170	45	191	20	67
84.	503	3	122	125	70	45	54	28	166	53	196	20	70
85.	538	5	137	134	64	41	54	28	162	47	208	21	72
86.	611	5	141	132	71	49	51	29	175	55	200	21	75
87.	536	5	135	135	64	39	50	28	166	45	193	20	73
88.	546	5	134	132	65	43	50	28	172	48	192	20	73
89.	411	3	132	124	65	35	43	26	152	46	174	19	68
90.	418	3	121	122	53	34	41	26	150	50	175	19	69
91.	466	5	127	126	76	46	46	27	162	46	177	19	69
92.	523	5	126	125	65	41	49	28	153	49	203	20	71
93.	447	3	125	127	64	36	44	26	154	44	184	19	67
94.	535	3	128	138	65	43	50	28	161	49	193	20	69
95.	604	5	125	128	66	45	54	28	165	51	205	21	74
96.	536	3	135	129	62	45	40	27	145	46	171	20	70
97.	533	3	131	132	61	45	48	28	167	48	197	20	71
98.	341	3	120	125	55	35	42	26	131	44	166	19	68

99	383	3	122	128	62	38	41	27	152	44	171	19	69
100	409	3	120	121	60	36	48	27	154	47	173	19	69

