

TESIS

**KOMPARASI ALGORITMA MACHINE LEARNING DENGAN
HYPERPARAMETER TUNING UNTUK KLASIFIKASI MULTI-LABEL
STATUS GIZI BALITA DI KABUPATEN MAJALENGKA**



Disusun oleh:

Nama : Fajar Maula Hidayat
NIM : 22.55.1213
Konsentrasi : Business Intelligence

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

2024

TESIS

**KOMPARASI ALGORITMA MACHINE LEARNING DENGAN
HYPERPARAMETER TUNING UNTUK KLASIFIKASI MULTI-LABEL
STATUS GIZI BALITA DI KABUPATEN MAJALENGKA**

**COMPARISON OF MACHINE LEARNING ALGORITHM WITH
HYPERPARAMETER TUNING FOR MULTI-LABEL CLASSIFICATION OF
NUTRITIONAL STATUS OF TODDLERS IN MAJALENGKA REGENCY**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Fajar Maula Hidayat
NIM : 22.55.1213
Konsentrasi : Business Intelligence

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

2024

HALAMAN PENGESAHAN

**KOMPARASI ALGORITMA MACHINE LEARNING DENGAN
HYPERPARAMETER TUNING UNTUK KLASIFIKASI MULTI-LABEL
STATUS GIZI BALITA DI KABUPATEN MAJALENGKA**

**COMPARISON OF MACHINE LEARNING ALGORITHM WITH
HYPERPARAMETER TUNING FOR MULTI-LABEL CLASSIFICATION OF
NUTRITIONAL STATUS OF TODDLERS IN MAJALENGKA REGENCY**

Dipersiapkan dan Disusun oleh

Fajar Maula Hidayat

22.55.1213

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Jumat, 5 Juli 2024

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

Yogyakarta, 5 Juli 2024

Rektor

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

HALAMAN PERSETUJUAN

KOMPARASI ALGORITMA MACHINE LEARNING DENGAN HYPERPARAMETER TUNING UNTUK KLASIFIKASI MULTI-LABEL STATUS GIZI BALITA DI KABUPATEN MAJALENGKA

COMPARISON OF MACHINE LEARNING ALGORITHM WITH HYPERPARAMETER TUNING FOR MULTI-LABEL CLASSIFICATION OF NUTRITIONAL STATUS OF TODDLERS IN MAJALENGKA REGENCY

Dipersiapkan dan Disusun oleh

Fajar Maula Hidayat

22.55.1213

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Jumat, 5 Juli 2024

Pembimbing Utama

Anggota Tim Penguji

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

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

Pembimbing Pendamping

Dr. Ferry Wahyu Wibowo, S.Si., M.Sc.
NIK. 190302235

Ainul Yaqin, M.Kom.
NIK. 190302255

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

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

Yogyakarta, 5 Juli 2024

Direktur Program Pascasarjana

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

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertanda tangan di bawah ini,

Nama mahasiswa : Fajar Maula Hidayat
NIM : 22.55.1213
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:
Komparasi Algoritma *Machine Learning* dengan *Hyperparameter Tuning* untuk Klasifikasi Multi-Label Status Gizi Balita di Kabupaten Majalengka

Dosen Pembimbing Utama : Prof. Dr. Kusriani, M.Kom,
Dosen Pembimbing Pendamping : Ainul Yaqin, M.Kom.

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, 5 Juli 2024
Yang Menyatakan,



Fajar Maula Hidayat

HALAMAN PERSEMBAHAN

Segala puji bagi Allah SWT atas segala nikmat, rahmat, dan karunia-Nya. Tesis ini kupersembahkan sebagai bentuk rasa syukur kepada-Nya. Kepada istriku tercinta (Ai Siti Hapidah, S.Ak.), anaku tersayang (Aifa Arunika Maula), serta orang tua terkasih atas cinta, dukungan, dan do'a yang tiada henti. Juga kepada dosen pembimbing dan seluruh pihak yang telah memberikan bantuan dan dukungan dalam penelitian ini. Terima kasih atas segala kebaikan dan kerjasama yang telah diberikan, semoga tesis ini bermanfaat bagi ilmu pengetahuan.



HALAMAN MOTTO

"Sesungguhnya beserta kesulitan itu pasti ada kemudahan."

– (QS. Al-Insyirah: 6)

*"Pengetahuan yang baik adalah yang memberikan manfaat,
bukan hanya diingat."*

– Imam Syafi'i



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan tesis ini dengan baik. Penyusunan tesis ini tidak lepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom., Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta serta sebagai pembimbing utama penelitian tesis ini, yang telah memberikan bimbingan, arahan, serta kesabaran selama proses penelitian ini.
3. Bapak Ainul Yaqin, M.Kom., sebagai pembimbing pendamping, atas masukan dan saran berharga sehingga hasil akhir dari penelitian ini menjadi semakin baik.
4. Bapak Dhani Ariatmanto, M.Kom., Ph.D., sebagai anggota tim penguji seminar proposal tesis, seminar hasil proposal tesis, dan ujian tesis, atas koreksi dan masukan yang membangun untuk penelitian tesis ini.
5. Bapak Dr. Ferry Wahyu Wibowo, S.St., M.Sc., sebagai anggota tim penguji seminar proposal tesis, seminar hasil proposal tesis, dan ujian tesis, atas koreksi dan sarannya untuk penelitian ini.
6. Seluruh staf admisi S2 Informatika, atas segala bantuan dan pelayanan yang diberikan selama masa studi di Universitas AMIKOM Yogyakarta.

7. Istri tercinta (Ai Siti Hapidah, S.Ak.) dan anaku tersayang (Aifa Arunika Maula), serta seluruh keluarga besar atas segala dukungan dan do'anya.
8. Rekan-rekan seperjuangan program studi S2 Informatika angkatan 7 tahun 2022, atas semangat kebersamaan dan dukungan yang tak ternilai perkuliahan jarak jauh.

Penulis berharap hasil penelitian ini dapat memberikan kontribusi dan manfaat yang baik bagi pengembangan ilmu pengetahuan. Akhir kata, penulis mohon maaf atas segala kekurangan dan kesalahan yang ada dalam penulisan tesis ini.

Yogyakarta, 5 Juli 2024

Penulis



DAFTAR ISI

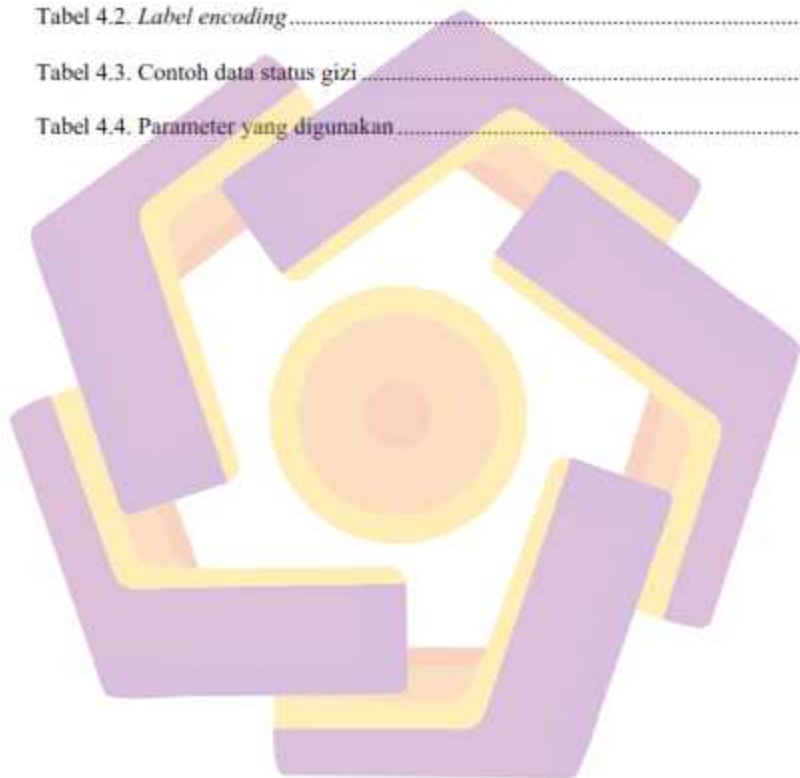
HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR ISTILAH.....	xviii
INTISARI.....	xix
<i>ABSTRACT</i>	xx
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	7
1.3. Batasan Masalah.....	8
1.4. Tujuan Penelitian.....	8
1.5. Manfaat Penelitian.....	9
BAB II TINJAUAN PUSTAKA.....	11
2.1. Tinjauan Pustaka.....	11

2.2. Keaslian Penelitian.....	15
2.3. Landasan Teori.....	19
2.3.1. Status Gizi Balita (<i>stunting</i> , <i>wasting</i> , dan <i>underweight</i>).....	19
2.3.2. <i>Knowledge Discovery in Databases (KDD)</i>	19
2.3.3. Proses <i>Knowledge Discovery in Database</i>	20
2.3.4. <i>Machine Learning</i>	21
2.3.5. Klasifikasi.....	22
2.3.6. Klasifikasi Multi-Label.....	23
2.3.7. <i>Naïve Bayes</i>	24
2.3.8. <i>Logistic Regression</i>	25
2.3.9. <i>Random Forest</i>	26
2.3.10. <i>Support Vector Classifier</i>	27
2.3.11. <i>Decision Tree</i>	28
2.3.12. <i>K-Nearest Neighbors</i>	30
2.3.13. <i>Confusion Matrix</i>	30
BAB III METODE PENELITIAN.....	33
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	33
3.1.1. Jenis Penelitian.....	33
3.1.2. Sifat Penelitian.....	33
3.1.3. Pendekatan Penelitian.....	34

3.2. Metode Pengumpulan Data.....	34
3.3. Metode Analisis Data.....	36
3.4. Alur Penelitian.....	38
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	41
4.1. <i>Data Understanding</i>	41
4.2. <i>Data Cleaning</i>	47
4.3. <i>Exploratory Data Analysis</i>	53
4.4. <i>Data Preparation</i>	59
4.5. <i>Modeling Machine Learning</i>	62
4.6. Evaluasi Model.....	78
4.7. <i>Testing</i>	97
4.8. Perbandingan Model.....	101
BAB V PENUTUP.....	104
5.1. Kesimpulan.....	104
5.2. Saran.....	105
DAFTAR PUSTAKA.....	106
LAMPIRAN.....	113

DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan posisi penelitian.....	15
Tabel 4.1. Deskripsi <i>dataset</i>	42
Tabel 4.2. <i>Label encoding</i>	60
Tabel 4.3. Contoh data status gizi.....	63
Tabel 4.4. Parameter yang digunakan.....	64



DAFTAR GAMBAR

Gambar 1.1. Prevalensi balita <i>stunting</i>	2
Gambar 1.2. Prevalensi balita <i>wasting</i>	2
Gambar 1.3. Prevalensi balita <i>underweight</i>	3
Gambar 2.1. Proses <i>knowledge discovery database (KDD)</i>	20
Gambar 2.2. Tahapan membangun model data latih (<i>data training</i>)	23
Gambar 2.3. Tahapan menggunakan model data uji (<i>data test</i>)	23
Gambar 2.4. <i>Decision tree</i>	29
Gambar 2.5. Tabel <i>Confusion matrix</i>	31
Gambar 3.1. Alur Penelitian	38
Gambar 4.1. <i>Dataset</i> status gizi	41
Gambar 4.2. Tipe data status gizi	43
Gambar 4.3. Distribusi usia balita	44
Gambar 4.4. Distribusi jenis kelamin balita	44
Gambar 4.5. Distribusi status gizi <i>stunting</i>	45
Gambar 4.6. Distribusi status gizi <i>wasting</i>	46
Gambar 4.7. Distribusi status gizi <i>underweight</i>	46
Gambar 4.8. Nilai yang hilang pada setiap atribut	47
Gambar 4.9. Data bernilai nol (0)	48
Gambar 4.10. Nilai yang hilang setelah <i>cleaning data</i>	49
Gambar 4.11. Data bernilai nol setelah <i>cleaning data</i>	49
Gambar 4.12. Konversi tipe data <i>object</i> ke <i>float64</i>	50

Gambar 4.13. <i>Box plot</i> berat badan lahir.....	50
Gambar 4.14. <i>Box plot</i> panjang atau tinggi badan lahir.....	51
Gambar 4.15. <i>Box plot</i> berat badan bayi.....	52
Gambar 4.16. <i>Box plot</i> tinggi badan bayi.....	52
Gambar 4.17. <i>Box plot</i> lingkaran lengan atas.....	53
Gambar 4.18. Distribusi usia balita.....	54
Gambar 4.19. Distribusi berat badan balita.....	55
Gambar 4.20. Distribusi tinggi badan balita.....	55
Gambar 4.21. Distribusi <i>stunting</i> berdasarkan jenis kelamin.....	56
Gambar 4.22. Distribusi <i>wasting</i> berdasarkan jenis kelamin.....	57
Gambar 4.23. Distribusi <i>underweight</i> berdasarkan jenis kelamin.....	57
Gambar 4.24. <i>Scatter plot</i> berat dan tinggi badan balita.....	58
Gambar 4.25. <i>Correlation matrix</i>	59
Gambar 4.26. Hasil <i>label encoding</i>	60
Gambar 4.27. <i>Feature importance</i>	61
Gambar 4.28. <i>Split</i> data latih dan data uji.....	62
Gambar 4.29. Hasil klasifikasi model <i>naïve bayes</i> label tb/u (<i>stunting</i>).....	65
Gambar 4.30. Hasil klasifikasi model <i>naïve bayes</i> label bb/u (<i>wasting</i>).....	66
Gambar 4.31. Hasil klasifikasi model <i>naïve bayes</i> label bb/tb (<i>underweight</i>).....	66
Gambar 4.32. Hasil klasifikasi model <i>logistic regression</i> label tb/u (<i>stunting</i>)....	67
Gambar 4.33. Hasil klasifikasi model <i>logistic regression</i> label bb/u (<i>wasting</i>) ...	68
Gambar 4.34. Hasil klasifikasi model <i>log regression</i> label bb/tb (<i>underweight</i>) .	68
Gambar 4.35. Hasil klasifikasi model <i>random forest</i> label tb/u (<i>stunting</i>).....	69

Gambar 4.36. Hasil klasifikasi model <i>random forest</i> label bb/u (<i>wasting</i>)	70
Gambar 4.37. Hasil klasifikasi model <i>random forest</i> label bb/tb (<i>underweight</i>)..	71
Gambar 4.38. Hasil klasifikasi model <i>svc</i> label tb/u (<i>stunting</i>)	72
Gambar 4.39. Hasil klasifikasi model <i>svc</i> label bb/u (<i>wasting</i>).....	72
Gambar 4.40. Hasil klasifikasi model <i>svc</i> label bb/tb (<i>underweight</i>)	73
Gambar 4.41. Hasil klasifikasi model <i>decision tree</i> label tb/u (<i>stunting</i>).....	74
Gambar 4.42. Hasil klasifikasi model <i>decision tree</i> label bb/u (<i>wasting</i>)	75
Gambar 4.43. Hasil klasifikasi model <i>decision tree</i> label bb/tb (<i>underweight</i>)....	75
Gambar 4.44. Hasil klasifikasi model <i>k-nearest neighbors</i> label tb/u (<i>stunting</i>)..	76
Gambar 4.45. Hasil klasifikasi model <i>k-nearest neighbors</i> label bb/u (<i>wasting</i>) .	77
Gambar 4.46. Hasil klasifikasi model <i>decision tree</i> label bb/tb (<i>underweight</i>)....	77
Gambar 4.47. <i>Confusion matrix stunting</i> (tb/u) model <i>naïve bayes</i>	79
Gambar 4.48. <i>Confusion matrix wasting</i> (bb/u) model <i>naïve bayes</i>	80
Gambar 4.49. <i>Confusion matrix underweight</i> (bb/tb) model <i>naïve bayes</i>	81
Gambar 4.50. <i>Confusion matrix stunting</i> (tb/u) model <i>logistic regression</i>	83
Gambar 4.51. <i>Confusion matrix wasting</i> (bb/u) model <i>logistic regression</i>	84
Gambar 4.52. <i>Confusion matrix underweight</i> (bb/tb) model <i>logistic regression</i> .	85
Gambar 4.53. <i>Confusion matrix stunting</i> (tb/u) model <i>random forest</i>	86
Gambar 4.54. <i>Confusion matrix wasting</i> (bb/u) model <i>random forest</i>	87
Gambar 4.55. <i>Confusion matrix underweight</i> (bb/tb) model <i>random forest</i>	88
Gambar 4.56. <i>Confusion matrix stunting</i> (tb/u) model <i>svc</i>	89
Gambar 4.57. <i>Confusion matrix wasting</i> (bb/u) model <i>svc</i>	90
Gambar 4.58. <i>Confusion matrix underweight</i> (bb/tb) model <i>svc</i>	91

Gambar 4.59. <i>Confusion matrix stunting (tb/u) model decision tree</i>	92
Gambar 4.60. <i>Confusion matrix wasting (bb/u) model decision tree</i>	93
Gambar 4.61. <i>Confusion matrix underweight (bb/tb) model decision tree</i>	94
Gambar 4.62. <i>Confusion matrix stunting (tb/u) model k-nearest neighbors</i>	95
Gambar 4.63. <i>Confusion matrix wasting (bb/u) model k-nearest neighbors</i>	96
Gambar 4.64. <i>Confusion matrix underweight (bb/tb) model k-nearest neighbors</i>	97
Gambar 4.65. Hasil pengujian model <i>naïve bayes</i>	97
Gambar 4.66. Jumlah nilai berbeda model <i>naïve bayes</i>	98
Gambar 4.67. Hasil pengujian model <i>logistic regression</i>	98
Gambar 4.68. Jumlah nilai berbeda model <i>logistic regression</i>	98
Gambar 4.69. Hasil pengujian model <i>random forest</i>	99
Gambar 4.70. Jumlah nilai berbeda model <i>random forest</i>	99
Gambar 4.71. Hasil pengujian model <i>svc</i>	99
Gambar 4.72. Jumlah nilai berbeda model <i>svc</i>	100
Gambar 4.73. Hasil pengujian model <i>decision tree</i>	100
Gambar 4.74. Jumlah nilai berbeda model <i>decision tree</i>	100
Gambar 4.75. Hasil pengujian model <i>k-nearest neighbors</i>	101
Gambar 4.76. Jumlah nilai berbeda model <i>k-nearest neighbors</i>	101
Gambar 4.77. Perbandingan model akurasi terbaik	102
Gambar 4.78. Perbandingan akurasi model <i>machine learning</i> setiap label	103

DAFTAR ISTILAH

jk (Jenis Kelamin)

Merujuk pada kategori biologis laki-laki atau perempuan.

bb_lahir (Berat Badan Lahir)

Berat badan bayi saat lahir, biasanya diukur dalam kilogram.

tb_lahir (Tinggi Badan Lahir)

Panjang atau tinggi badan bayi saat lahir, biasanya diukur dalam sentimeter.

lila (Lingkar Lengan Atas)

Pengukuran lingkar lengan atas, biasanya diukur dalam sentimeter.

tb/u (Tinggi Badan / Usia)

z-score yang mengukur tinggi badan balita dibandingkan dengan standar tinggi badan balita pada usia yang sama. Digunakan untuk mengevaluasi apakah balita mengalami *stunting*.

bb/u (Berat Badan / Usia)

z-score yang mengukur berat badan balita dibandingkan dengan standar berat badan balita pada usia yang sama. Digunakan untuk mengevaluasi apakah balita mengalami *wasting*.

bb/tb (Berat Badan / Tinggi Badan)

z-score yang mengukur berat badan balita dibandingkan dengan tinggi badan. Digunakan untuk mengevaluasi apakah balita mengalami *overweight*.

z-score

Ukuran standar deviasi dari nilai data yang menunjukkan seberapa jauh nilai tersebut dari rata-rata.

outlier

Data yang sangat berbeda dari data lainnya, dapat mempengaruhi hasil analisis.

INTISARI

Status gizi balita merupakan indikator penting dalam pemantauan kesehatan masyarakat yang memerlukan pendekatan efektif dalam pencegahan dan penanganannya. Penelitian ini bertujuan untuk membandingkan kinerja beberapa algoritma *machine learning* dengan *hyperparameter tuning* untuk klasifikasi multi-label status gizi balita di Kabupaten Majalengka. Algoritma yang dibandingkan meliputi *naïve bayes*, *logistic regression*, *random forest*, *svc*, *decision tree* dan *k-nearest neighbors*, dengan masing-masing algoritma melalui penyetelan *hyperparameter* untuk optimalisasi kinerja model. Data yang digunakan mencakup berbagai indikator kesehatan balita yang diperoleh dari Dinas Kesehatan Kabupaten Majalengka. Hasil penelitian menunjukkan bahwa algoritma *Support Vector Classification (SVC)* mencapai akurasi tertinggi dalam klasifikasi multi-label status gizi balita, dengan akurasi sebesar 94,28% untuk *tb/u (stunting)*, 94,28% untuk *bb/u (wasting)*, dan 95,12% untuk *bb/tb (underweight)*. Hasil ini menunjukkan bahwa *SVC* merupakan algoritma yang paling efektif dalam mengklasifikasikan status gizi balita, sehingga dapat menjadi pilihan utama dalam menangani masalah status gizi pada balita, yang merupakan tantangan penting dalam bidang kesehatan masyarakat.

Kata kunci: status gizi balita, klasifikasi multi-label, *machine learning*, *hyperparameter tuning*

ABSTRACT

Nutritional status of toddlers is a crucial indicator in public health monitoring, requiring effective approaches for prevention and management. This study aims to compare the performance of several machine learning algorithms with hyperparameter tuning for multi-label classification of toddler nutritional status in Majalengka Regency. The algorithms compared include naïve bayes, logistic regression, random forest, svc, decision tree and k-nearest neighbors, each undergoing hyperparameter tuning to optimize model performance. The data used utilized encompasses various health indicators of toddlers obtained from the Majalengka District Health Office. The result show that the Support Vector Classification (SVC) algorithm achieved the highest accuracy in multi-label classification toddler nutritional status, with an accuracy of 94.28% for tb/u (stunting), 94.28% for bb/u (wasting), and 95.12% for bb/tb (underweight). These findings indicate that SVC is the most effective algorithm for classifying toddler nutritional status, making it a primary choice in addressing nutritional status issues in toddlers, a significant challenge in public health.

Keyword: toddler nutritional status, multi-label classification, machine learning, hyperparameter tuning



BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Kementerian Kesehatan mengumumkan hasil Survei Status Gizi Indonesia (SSGI) yang menunjukkan penurunan prevalensi *stunting* dari 24,4% pada tahun 2021 menjadi 21,6% pada tahun 2022, menurut hasil Rapat Kerja Nasional BKKBN pada Rabu, 25 Januari 2023. Menurut standar WHO, prevalensi *stunting* seharusnya di bawah 20%. Prevalensi *wasting* meningkat dari 7,1% pada tahun 2021 menjadi 7,7% pada tahun 2022. Prevalensi *underweight* juga meningkat dari 17% pada tahun 2021 menjadi 17,1% pada tahun 2022. Sedangkan prevalensi *overweight* menunjukkan penurunan dari 3,8% pada tahun 2021 menjadi 3,5% pada tahun 2022. Pemerintah menargetkan penurunan *stunting* hingga mencapai 14% pada tahun 2024 (Kementerian Kesehatan RI, 2023).

Hasil SSGI di Provinsi Jawa Barat pada tahun 2022 menunjukkan prevalensi balita *stunting* 20,2%, balita *wasting* 6,0%, balita *underweight* 14,2%, dan balita *overweight* 3,8% (Kementerian Kesehatan RI, 2022). Kabupaten Majalengka adalah salah satu dari sepuluh kota dan kabupaten tertinggi di Provinsi Jawa Barat dalam hal status gizi, dengan prevalensi balita *stunting* 24,3%, balita *wasting* 7,7%, balita *underweight* 16,1%, dan balita *overweight* 2,5%.

Bulan Penimbangan Balita (BPB) adalah kegiatan yang dilakukan secara bersamaan oleh posyandu pada bulan Agustus setiap tahunnya pada balita berusia antara 0 sampai 60 bulan dengan menimbang berat badan dan mengukur panjang

atau tinggi badan pada balita. Data Rekapitulasi Hasil BPB di Kabupaten Majalengka pada tahun 2022, yang diperoleh dari Dinas Kesehatan Kabupaten Majalengka, menunjukkan bahwa kecamatan Sumberjaya memiliki banyak balita *stunting* dan balita *underweight* tertinggi, seperti yang ditunjukkan pada Gambar 1.1 dan Gambar 1.3, sementara kecamatan Salagedang memiliki banyak balita *wasting* tertinggi, seperti yang ditunjukkan pada Gambar 1.2.



Gambar 1.1. Prevalensi balita *stunting*.



Gambar 1.2. Prevalensi balita *wasting*.



Gambar 1.3. Prevalensi balita *underweight*

Antropometri adalah metode langsung untuk menilai status gizi seseorang, terutama masalah kekurangan energi dan protein (KEP). Faktor genetik dan lingkungan, seperti konsumsi makanan dan kondisi kesehatan, mempengaruhi antropometri (Aritonang, 2013). Meskipun memiliki beberapa keuntungan, seperti prosedur yang sederhana, aman, dan dapat dilakukan dalam sampel besar, dan penapisan kelompok rawan gizi, antropometri juga memiliki kelemahan. Kelemahannya termasuk ketidakmampuan untuk mengetahui status gizi dengan cepat, kemungkinan dipengaruhi oleh hal-hal yang tidak berkaitan dengan gizi, dan kesalahan pengukuran yang mengurangi ketepatan dan akurasi pengukuran. Menurut Istiany dan Rusilanti (Istiany & Rusilanti, 2013).

Untuk mengatasi masalah ini, teknik yang lebih canggih harus digunakan, seperti penggunaan *machine learning*. Dengan menggunakan *machine learning*, statistik matematika, dan kecerdasan buatan (AI), dapat melakukan klasifikasi kondisi status gizi balita dalam jangka panjang, termasuk *stunting*, *wasting*, dan *underweight*. Penggunaan *machine learning* memungkinkan untuk menemukan

pola yang tersembunyi dalam data gizi dan dapat memberikan pemahaman yang lebih mendalam tentang faktor-faktor yang mempengaruhi status gizi balita. *Machine learning* dapat membantu mendeteksi status gizi dengan lebih akurat dan tepat. Ini dapat membantu dalam pencegahan dan intervensi yang lebih efektif untuk meningkatkan kesehatan balita dan mencegah masalah status gizi di masa depan (Utomo & Mesran, 2020).

Menurut (Tarigan dkk., 2019), tahap klasifikasi adalah bagian penting dari proses analisis data. Tujuannya adalah untuk mengkategorikan item ke dalam kelompok atau kategori yang telah ditentukan sebelumnya. Dalam penelitian ini, metode klasifikasi *machine learning* akan digunakan. Algoritma ini memiliki kemampuan untuk memprediksi kapan suatu objek akan memuningkan menjadi anggota kelas tertentu (Darnila dkk., 2021).

Penelitian yang dilakukan oleh (R. Setiawan & Triayudi, 2022) mengenai klasifikasi status gizi balita, algoritma *naïve bayes* dan *k-nearest neighbors* digunakan sebagai metode analisis. Hasil penelitian menunjukkan bahwa algoritma *naïve bayes* mencapai tingkat akurasi sebesar 80,60%, sedangkan algoritma *k-nearest neighbors* mencapai tingkat akurasi sebesar 91,79%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma *k-nearest neighbors* lebih unggul dalam melakukan klasifikasi status gizi balita dibandingkan algoritma *naïve bayes*.

Penelitian lain yang dilakukan oleh (Afiatuddin dkk., 2024) telah membandingkan beberapa algoritma *machine learning* dalam klasifikasi kanker payudara. Hasil penelitian menunjukkan bahwa algoritma *logistic regression* berhasil mencatat hasil terbaik dengan tingkat akurasi 100,00%. Di posisi kedua,

algoritma *k-nearest neighbors* menunjukkan kinerja yang baik dengan tingkat akurasi mencapai 98.60%. Sementara, algoritma *random forest* mencapai tingkat akurasi sebesar 98.10%. Meskipun begitu, baik algoritma *decision tree* dan *naïve bayes* mencatat tingkat akurasi yang sama, yaitu 95.80%. Dengan demikian, dapat disimpulkan bahwa algoritma *logistic regression* menjadi pilihan yang paling unggul dalam mengklasifikasikan kanker payudara dibandingkan dengan algoritma lain yang diuji.

Selanjutnya penelitian yang dilakukan oleh (Rizquallah dkk., 2024) membahas perbandingan penerapan algoritma *c4.5*, *k-nearest neighbors*, dan *naïve bayes* untuk memprediksi keberlangsungan hidup pasien dengan gagal jantung. Hasil penelitian ini menunjukkan bahwa algoritma *c4.5* mencapai tingkat akurasi sebesar 77.91%, sedangkan algoritma *k-nearest neighbors* mencapai tingkat akurasi sebesar 63.54%. Sementara, algoritma *naïve bayes* memperoleh tingkat akurasi sebesar 76.59%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma *c4.5* menunjukkan kinerja yang paling unggul dalam klasifikasi keberlangsungan pasien gagal jantung dibandingkan dengan algoritma *k-nearest neighbors* dan *naïve bayes*.

Penelitian berikutnya oleh (Laela & Baihaqi, 2024) menganalisis klasifikasi penyakit *multiple sclerosis* dengan menggunakan algoritma *logistic regression* dan *support vector machine*. Hasil penelitian menunjukkan bahwa algoritma *support vector machine* mencapai tingkat akurasi sebesar 88.33%, sementara algoritma *logistic regression* mencapai tingkat akurasi sebesar 86.67%. Berdasarkan hasil ini, dapat disimpulkan bahwa algoritma *support vector machine* lebih unggul dalam

klasifikasi penyakit *multiple sclerosis* dibandingkan dengan algoritma *logistic regression*.

Penelitian lain yang dilakukan oleh (A. Setiawan dkk., 2024) membahas perbandingan algoritma *k-nearest neighbors*, *support vector machine*, dan *decision tree* dalam klasifikasi penyakit stroke. Hasil penelitian menunjukkan bahwa algoritma *k-nearest neighbors* mencapai tingkat akurasi sebesar 94%, algoritma *support vector machine* mencapai tingkat akurasi sebesar 95%, dan algoritma *decision tree* memperoleh tingkat akurasi sebesar 92%. Kesimpulan dari penelitian ini adalah bahwa algoritma *support vector machine* memiliki kinerja terbaik dalam mengklasifikasi penyakit stroke dibandingkan dengan algoritma *k-nearest neighbors* dan *decision tree*.

Kemudian penelitian yang dilakukan oleh (Aditya dkk., 2024) mengenai prediksi penyakit hipertensi menggunakan metode *decision tree* dan *random forest*. Hasil penelitian menunjukkan bahwa kedua algoritma tersebut memperoleh tingkat akurasi yang tinggi, yaitu 100%. Dapat disimpulkan bahwa, baik *decision tree* atau *random forest* memberikan prediksi yang sangat akurat untuk penyakit hipertensi.

Berdasarkan uraian singkat terhadap penelitian terdahulu yang relevan dengan penelitian saat ini, dapat disimpulkan bahwa terdapat variasi dalam kinerja berbagai algoritma *machine learning* dalam berbagai konteks kesehatan, seperti klasifikasi status gizi balita, keberlangsungan pasien dengan penyakit tertentu, dan prediksi penyakit. Berbagai algoritma seperti *naïve bayes*, *logistic regression*, *random forest*, *svc*, *decision tree*, dan *k-nearest neighbors* telah diuji dalam berbagai penelitian ini.

Namun, terdapat perbedaan dalam tingkat akurasi dan performa antara algoritma-algoritma tersebut. Beberapa peneliti menunjukkan bahwa beberapa algoritma memiliki tingkat akurasi yang lebih tinggi daripada yang lain dalam konteks tertentu. Misalnya, dalam penelitian tertentu, *decision tree* memiliki akurasi tertinggi, sedangkan pada penelitian lain, *support vector machine* atau *random forest* lebih unggul.

Hal ini menunjukkan bahwa terdapat ruang untuk eksplorasi lebih lanjut dalam memilih algoritma yang paling sesuai untuk konteks penelitian yang dilakukan saat ini. Dengan memahami hasil-hasil penelitian terdahulu, penelitian memiliki dasar yang kuat untuk mengeksplorasi kinerja berbagai algoritma klasifikasi status gizi balita. Oleh karena itu, penelitian ini penting untuk mengidentifikasi algoritma *machine learning* yang paling efektif dan akurat dalam mengklasifikasi status gizi balita. Diharapkan penelitian ini akan menghasilkan temuan yang relevan dan akurat untuk memberikan wawasan yang berharga dalam upaya pencegahan dan penanganan masalah gizi pada balita, terutama di Kabupaten Majalengka.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah, rumusan masalah yang akan dibahas pada penelitian ini adalah sebagai berikut:

- a. Algoritma *machine learning* apa yang memiliki tingkat akurasi tertinggi dalam mengklasifikasi multi-label status gizi balita?

- b. Kombinasi *hyperparameter* apa yang menghasilkan akurasi terbaik pada algoritma dalam mengklasifikasi multi-label status gizi balita?

1.3. Batasan Masalah

Untuk menjaga arah dan fokus pada penelitian ini, maka dibuatlah batasan masalah sebagai berikut:

- a. Data yang digunakan untuk penelitian ini berasal dari Rekapitulasi Hasil Bulan Penimbangan Balita (BPB) tahun 2022 di Kabupaten Majalengka, yang diperoleh dari Dinas Kesehatan Kabupaten Majalengka.
- b. Penelitian ini menggunakan 3.504 data dari Puskesmas Sumberjaya, yang memiliki jumlah kasus status gizi tertinggi di Kabupaten Majalengka.
- c. Model *machine learning* dalam penelitian ini hanya menggunakan algoritma *naïve bayes*, *logistic regression*, *random forest*, *svc*, *decision tree* dan *k-nearest neighbors*.
- d. Lingkup penelitian ini terbatas pada klasifikasi multi-label *stunting*, *wasting*, dan *underweight* pada balita, dan tidak membahas masalah status gizi pada orang dewasa atau lansia.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah, penelitian ini memiliki tujuan diantaranya:

- a. Mengidentifikasi algoritma *machine learning* yang memiliki tingkat akurasi tertinggi dalam mengklasifikasi multi-label status gizi balita?

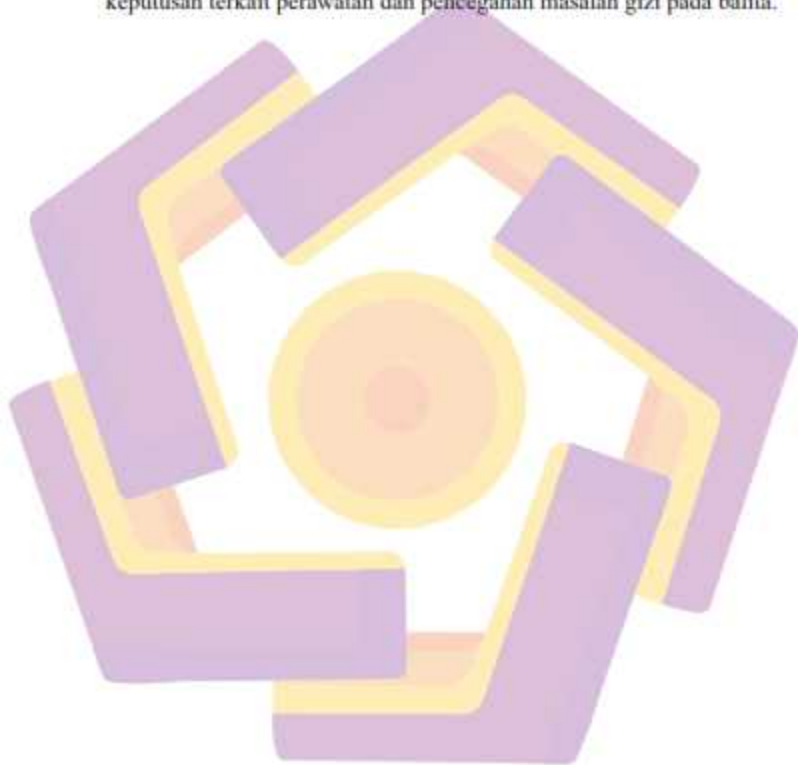
- b. Menentukan kombinasi *hyperparameter* yang menghasilkan akurasi terbaik pada algoritma dalam mengklasifikasi multi-label status gizi balita?

1.5. Manfaat Penelitian

Manfaat yang akan diperoleh dari penelitian ini adalah:

- a. Kontribusi Ilmiah: Penelitian ini memberikan kontribusi dalam pengembangan metode klasifikasi multi-label dalam bidang kesehatan, khususnya pemodelan status gizi balita. Hasil penelitian dapat digunakan sebagai referensi untuk penelitian lebih lanjut dalam pengembangan algoritma *machine learning* yang lebih efektif dan efisien dalam klasifikasi multi-label. Penelitian ini juga dapat memperkaya literatur mengenai penggunaan *hyperparameter tuning* untuk meningkatkan akurasi model dalam aplikasi kesehatan.
- b. Lingkungan Pengembangan: Bagi pengembang aplikasi kesehatan, penelitian ini memberikan panduan dalam memilih algoritma dan konfigurasi *hyperparameter* yang optimal untuk klasifikasi status gizi balita. Penelitian ini membantu dalam meningkatkan keandalan dan akurasi aplikasi berbasis *machine learning* yang digunakan dalam monitoring dan evaluasi status gizi di masyarakat. Hasil penelitian dapat digunakan sebagai dasar dalam pengembangan *tools* atau perangkat lunak yang secara otomatis dapat menyesuaikan *hyperparameter* untuk meningkatkan performa klasifikasi.

- c. Pengguna: Penelitian ini bermanfaat bagi tenaga kesehatan dalam mengidentifikasi secara lebih akurat status gizi balita, sehingga intervensi yang diberikan dapat lebih tepat sasaran. Dengan model yang lebih akurat, pengguna di lapangan dapat lebih percaya diri dalam pengambilan keputusan terkait perawatan dan pencegahan masalah gizi pada balita.



BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Beberapa penelitian-penelitian sebelumnya dijadikan sebagai bahan acuan dan tinjauan pustaka untuk penelitian yang dilakukan saat ini diantaranya adalah:

Penelitian oleh (Rahman dkk., 2021) melakukan penelitian malnutrisi pada balita di Bangladesh. Mereka menerapkan algoritma *machine learning* untuk mengidentifikasi dan memprediksi faktor risiko malnutrisi, seperti *stunting*, *wasting*, dan *underweight*. Sumber data yang digunakan diperoleh dari Survei Demografi dan Kesehatan Bangladesh tahun 2014, yang mencakup populasi 7.079 balita dengan 13 variabel faktor yang relevan. Analisis dilakukan dengan memanfaatkan *logistic regression* untuk mengidentifikasi potensi risiko malnutrisi. Selain itu, tiga pendekatan *machine learning*, yakni *support vector machine*, *random forest*, dan *logistic regression*, diterapkan untuk memprediksi malnutrisi dan performa algoritma *machine learning* ini dinilai berdasarkan tingkat akurasi. Hasil penelitian menunjukkan bahwa prevalensi *stunting*, *wasting*, dan *underweight* masing-masing mencapai 35,4%, 15,4%, dan 32,8%. Selain itu, *logistic regression* mampu mengidentifikasi lima faktor risiko *stunting* dan *underweight*, serta empat faktor *wasting*. Algoritma *random forest* adalah model klasifikasi terbaik yang efektif, mencapai tingkat akurasi tertinggi sebesar 88,3% untuk *stunting*, 87,7% untuk *wasting*, dan 85,7% untuk *underweight*.

Selanjutnya penelitian yang dilakukan oleh (Chilyabanyama dkk., 2022) melakukan penelitian *stunting* pada balita di Zambia. Mereka menerapkan algoritma *machine learning*, termasuk *logistic regression*, *random forest*, *support vector machine*, *XGboost*, dan *naïve bayes*, untuk memprediksi kemungkinan *stunting* pada balita. Data nutrisi yang digunakan berasal dari Survei Kesehatan Demografi Zambia (ZDHS) tahun 2018, yang dilaksanakan oleh Badan Statistik Zambia bekerja sama dengan USAID. Analisis dilakukan dengan mengukur *accuracy*, *precision*, *recall*, dan *f1-score* dari masing-masing algoritma *machine learning*. Tiga belas dari lima puluh delapan fitur dipilih untuk dimasukkan ke dalam model *random forest*. Hasil menunjukkan bahwa *random forest* menjadi algoritma yang paling akurat, mencapai tingkat akurasi sebesar 79% pada pengujian dan 61,6% pada data pelatihan, sedangkan *naïve bayes* menunjukkan algoritma dengan kinerja terburuk dalam memprediksi *stunting* pada balita di Zambia menggunakan *dataset* ZDHS tahun 2018.

Selanjutnya penelitian yang dilakukan oleh (Yanto dkk., 2023) melakukan analisis klasifikasi untuk mengetahui status gizi dengan meningkatkan kinerja *machine learning*. Data yang digunakan berasal dari kasus gizi anak yang terdiri dari 576 pasien yang bersumber dari Rumah Sakit Umum Daerah (RSUD) M. Djamil Padang. Data tersebut terbagi menjadi 417 data latih dan 159 data uji. Proses analisis klasifikasi *machine learning* dilakukan menggunakan metode jaringan syaraf tiruan dengan algoritma *multilayer perceptron*. Pengoptimalan kinerja *machine learning* dilakukan dengan memanfaatkan metode korelasi *pearson* untuk mendapatkan pola analisis klasifikasi yang optimal. Hasil pengujian menunjukkan

bahwa metode korelasi *pearson* dapat memberikan pola analisis yang tepat dan akurat, dengan tingkat akurasi mencapai 95%.

Selanjutnya penelitian yang dilakukan oleh (Yunus dkk., 2023) melakukan klasifikasi *stunting* pada balita berusia 0-60 bulan dengan menerapkan metode *decision tree (C4.5)* berdasarkan perhitungan *z-score*. Data yang digunakan berasal dari survei kejadian *stunting* di Kabupaten Gorontalo tahun 2018, terdiri dari 224 sampel, yang mencakup 4 atribut fitur dan 1 atribut label, yaitu jenis kelamin, usia, berat badan, tinggi badan, dan status gizi. Hasil penelitian menunjukkan bahwa *decision tree (C4.5)* mengidentifikasi usia sebagai fitur yang paling berpengaruh dalam klasifikasi *stunting*, dengan *gain ratio* tertinggi 0.185016337. Evaluasi model menggunakan *confusion matrix* menghasilkan akurasi tertinggi sebesar 61,82% dan nilai AUC sebesar 0,584.

Selanjutnya penelitian yang dilakukan oleh (Lonang, Yudhana, & Kunta Biddinika, 2023) melakukan evaluasi komparatif lima algoritma *machine learning*, yakni *logistic regression*, *decision tree*, *random forest*, *k-nearest neighbors*, dan *support vector machine* dalam mengklasifikasikan *stunting* pada balita. *Dataset* yang digunakan berasal dari Elektronik Pencatatan dan Pelaporan Gizi Balita Berbasis Masyarakat (EPPGBM) Puskesmas Joggat, Lombok Tengah, yang terdiri dari 1.295 sampel data, termasuk 1.000 balita yang tidak terindikasi *stunting* dan 295 balita yang terindikasi *stunting*. *Dataset* terdiri 5 atribut fitur, seperti jenis kelamin, usia, tinggi badan, berat badan, dan atribut label berupa kelas biner. Dari hasil pengujian dan analisis perbandingan algoritma *machine learning*, *k-nearest neighbors* pada $k = 3$ mencapai akurasi tertinggi sebesar 94,85%, menunjukkan

kinerja terbaik dalam mengklasifikasikan *stunting* pada balita. Sementara itu, *decision tree* dan *random forest* juga memperoleh akurasi yang tinggi sebesar 94,34%. Di sisi lain, *logistic regression* menunjukkan akurasi yang lebih sedikit rendah sebesar 86,63%, dan *support vector machine* mencapai akurasi terendah sebesar 85,60%.

Penelitian lain yang dilakukan oleh (Ula dkk., 2022) melakukan klasifikasi status gizi anak dengan menerapkan model *decision tree*, yang kemudian hasil akhirnya dapat mengidentifikasi klasifikasi kerentanan gizi pada setiap pasien di RSUD Aceh Utara. *Dataset* yang digunakan terdiri dari 4 atribut fitur, yakni jenis kelamin, umur, berat badan, tinggi badan, dan 1 atribut label yang menunjukkan klasifikasi status gizi, termasuk gizi buruk, tidak gizi buruk, dan gizi lebih. Hasil pengujian penerapan *machine learning* menggunakan *decision tree* menunjukkan nilai rata-rata akurasi sebesar 81,25%, dengan menggunakan data training 40, 72, 146, 187, dan data testing 195 data.

2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian
 Komparasi Algoritma *Machine Learning* dengan *Hyperparameter Tuning* untuk Klasifikasi Multi-Label Status Gizi Balita di Kabupaten Majalengka

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	<i>Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach</i>	S. M. Jubaidur Rahman ¹ , N. A. M. Faisal Ahmed ² , Md. Menhajul Abedin ¹ , Benojir Ahammed ³ , Mohammad Ali ⁴ , Md. Jahnur Rahman ⁵ , Md. Maniruzzaman ^{1*}	Memfokuskan pada epidemiologi malnutrisi pada balita di Bangladesh dengan menerapkan algoritma <i>machine learning</i> untuk mengidentifikasi dan memprediksi mendeteksi faktor risiko malnutrisi (<i>stunted, wasted, dan underweight</i>).	Hasil penelitian menunjukkan prevalensi <i>stunting</i> 35,4%, <i>wasting</i> 15,4%, dan <i>underweight</i> 32,8%. <i>Logistic regression</i> berhasil mengidentifikasi lima faktor risiko untuk <i>stunting, wasting, dan underweight</i> , serta empat faktor <i>wasting</i> . <i>Random forest</i> terbukti menjadi algoritma yang paling efektif dengan tingkat akurasi tertinggi tertinggi 88,3% untuk <i>stunting</i> , 87,7% untuk <i>wasting</i> , dan 85,7% untuk <i>underweight</i> .	Penelitian ini menyarankan penggunaan lebih banyak variabel dan data longitudinal untuk meningkatkan akurasi. Salah satu kelemahannya adalah cakupan data yang terbatas pada satu tahun survei.	Penelitian saat ini menggunakan berbagai algoritma <i>machine learning</i> untuk klasifikasi multi-label status gizi pada balita, tetapi dengan fokus pada populasi yang berbeda-beda dan set data yang lebih baru. Algoritma yang digunakan lebih beragam, termasuk <i>naive bayes, logistic regression, random forest, svc, decision tree, dan k-nearest neighbors</i> .

Tabel 2.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2	<i>Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia</i>	Obvious Nchimunya Chiyabanyama ^{1,2,3} , Roma Chilengi ² , Michelo Simuyandi ² , Caroline C. Chisenga ² , Masuzyo Chirwa ² , Kalongo Hamusonde ² , Rakesh Kumar Saroj ² , Najeeha Talat Iqbal ² , Innocent Ngaruye ² , Samuel Bosomprah ^{2,4}	Meneliti <i>stunting</i> pada balita di Zambia, melatih dengan menerapkan algoritma <i>machine learning</i> untuk memprediksi kemungkinan <i>stunting</i> .	<i>Random forest</i> merupakan algoritma yang paling akurat dengan tingkat akurasi 79% pada pengujian dan 61,6% pada data pelatihan. <i>Naive bayes</i> memiliki kinerja terburuk dalam memprediksi <i>stunting</i> .	Saran untuk penelitian selanjutnya adalah meningkatkan pemilihan fitur dan penggunaan data yang lebih beragam. Kelemahan termasuk rendahnya akurasi pada data pelatihan dan keterbatasan pada jumlah fitur yang digunakan.	Penelitian saat ini juga memanfaatkan berbagai algoritma <i>machine learning</i> untuk klasifikasi multi-label status gizi pada balita, namun dengan data dan populasi yang berbeda. Penelitian ini menekankan pada pentingnya pemilihan fitur yang optimal dan penggunaan berbagai algoritma untuk mendapatkan hasil terbaik.
3	<i>Optimization of Machine Learning Classification Analysis of Malnutrition Cases in Children</i>	Musli Yanto ^{1,2} , Febri Hadri ² , Syafri Arlis ²	Melakukan analisis klasifikasi untuk mengetahui status gizi dengan meningkatkan kinerja <i>machine learning</i> . Proses analisis klasifikasi <i>machine learning</i> nantinya akan memanfaatkan kinerja metode jaringan syaraf tiruan (JST)	Metode korelasi <i>pearson</i> meningkatkan kinerja <i>machine learning</i> , menghasilkan tingkat akurasi sebesar 95% dalam analisis klasifikasi.	Saran untuk penelitian selanjutnya adalah eksplorasi lebih lanjut mengenai fitur-fitur lain yang mungkin berkontribusi terhadap peningkatan akurasi. Kelemahan termasuk terbatasnya jumlah data dan variasi fitur yang digunakan.	Penelitian saat ini juga fokus pada penggunaan algoritma <i>machine learning</i> untuk klasifikasi multi-label status gizi pada balita, dengan tujuan untuk meningkatkan akurasi dan efisiensi klasifikasi. Penelitian ini menekankan pentingnya pengoptimalan algoritma dan pemilihan fitur yang tepat.

Tabel 2.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
			dengan algoritma <i>Multilayer Perceptron</i> (MLP). Performa <i>machine learning</i> dapat dioptimalkan menggunakan metode korelasi <i>Pearson</i> (PC) untuk menghasilkan pola analisis klasifikasi yang optimal.			
4	<i>Classification of Stunting in Children Using the C4.5 Algorithm</i>	Muhajir Yunus ^{1*} , Muhammad Kunta Biddinika ² , Abdul Fadli ³	Mengklasifikasikan <i>stunting</i> pada anak usia 0-60 bulan dengan menerapkan metode <i>decision tree</i> (C4.5) berdasarkan perhitungan <i>z-score</i> .	<i>Decision tree</i> (c4.5) mengidentifikasi usia sebagai variabel yang paling berpengaruh dalam klasifikasi <i>stunting</i> , dengan gain ratio 0.185016337. akurasi model tertinggi diperoleh adalah 61,82%, dengan nilai AUC sebesar 0,584.	Saran untuk penelitian selanjutnya termasuk eksplorasi algoritma lain yang mungkin memberikan akurasi lebih tinggi. Kelemahan penelitian ini adalah akurasi yang relatif rendah dan nilai AUC yang tidak terlalu signifikan.	Penelitian saat ini juga menggunakan algoritma <i>machine learning</i> untuk klasifikasi multi-label status gizi pada balita. Fokus penelitian saat ini adalah pada peningkatan akurasi dan identifikasi variabel yang paling berpengaruh serupa dengan pendekatan yang digunakan Yunus, dkk. Dalam mengidentifikasi usia sebagai faktor penting.
5	<i>Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting</i>	Syahani Lonang ¹ , Anton Yudhana ^{2*} , Muhammad Kunta Biddinika ¹	Melakukan analisis komparatif lima algoritma <i>machine learning</i> , khususnya <i>logistic regression</i> ,	<i>K-nearest neighbors</i> pada $k=3$ mencapai akurasi tertinggi sebesar 94,85%, menunjukkan kinerja	Penelitian ini menunjukkan perlunya eksplorasi lebih lanjut pada parameter yang optimal untuk setiap	Penelitian saat ini juga berfokus pada komparasi algoritma <i>machine learning</i> untuk klasifikasi multi-label status gizi pada balita,

Tabel 2.1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
			<i>decision tree</i> , <i>random forest</i> , <i>k-nearest neighbors</i> , dan <i>support vector machine</i> , dalam konteks deteksi <i>stunting</i> pada balita.	terbaik dalam mengklasifikasikan <i>stunting</i> pada balita. <i>Decision tree</i> dan <i>random forest</i> juga memperoleh akurasi tinggi sebesar 94,34%. <i>Logistic regression</i> menunjukkan akurasi sebesar 86,63%, dan <i>support vector machine</i> mencapai akurasi terendah sebesar 85,60%.	algoritma guna meningkatkan akurasi klasifikasi. Kelemahan lain adalah variasi akurasi yang cukup besar antara algoritma yang diuji.	dengan tujuan mengidentifikasi model yang paling efektif dan akurat, mirip dengan pendekatan yang diambil oleh Lonang, dkk.
6	<i>Application Of Machine Learning In Determining The Classification Of Children's Nutrition With Decision Tree</i>	Mutammimul Ula ^{1*} , Ananda Faridhatul Ulva ² , Mauliza ³ , Muhammad Abdullah Ali ⁴ , Yunna Rilasmi Said ⁵	Mengklasifikasikan status gizi anak dengan menggunakan model <i>decision tree</i> untuk mengidentifikasi klasifikasi kerentanan gizi pada setiap pasien di RSUD Aceh Utara.	Hasil pengujian penerapan <i>machine learning</i> dengan model <i>decision tree</i> menunjukkan nilai rata-rata akurasi sebesar 81,25%, dengan menggunakan data <i>training</i> sebanyak 40, 72, 146, dan 187, serta data <i>testing</i> sebanyak 195 sampel.	Penelitian ini memberikan kontribusi dalam upaya identifikasi dan klasifikasi status gizi anak di RSUD Aceh Utara. Namun, akurasi yang diperoleh mungkin masih dapat ditingkatkan dengan model lainnya.	Penelitian saat ini juga menggunakan pendekatan <i>machine learning</i> , khususnya pada klasifikasi multi-label status gizi pada balita. Menganalisis dengan perbandingan kinerja algoritma yang dihasilkan, dan memiliki tujuan yang serupa dalam mengidentifikasi dan klasifikasi status gizi pada balita.

2.3. Landasan Teori

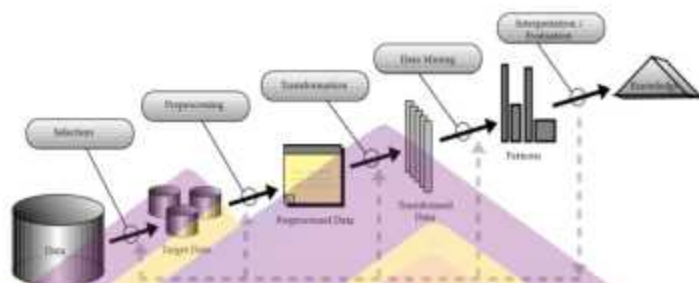
2.3.1. Status Gizi Balita (*stunting, wasting, dan underweight*)

Menurut (Supriasa dkk., 2016), status gizi merepresentasikan keseimbangan nutrisi dalam bentuk variabel tertentu atau representasi kecukupan nutrisi dalam bentuk variabel tertentu. Salah satu metode status gizi adalah melalui penggunaan pengukuran antropometri, yang dalam penelitian ini mencakup indeks tinggi badan menurut umur (tb/u), indeks berat badan menurut umur (bb/u), dan indeks berat badan menurut tinggi badan (bb/tb). Indeks tersebut memberikan indikasi adanya masalah gizi yang sifatnya kronis akibat kondisi yang berlangsung dalam jangka waktu yang lama, seperti kemiskinan, gaya hidup tidak sehat, dan asupan makanan kurang. Dampak dari kondisi ini bisa berupa pertumbuhan terhambat sehingga mengakibatkan anak menjadi pendek, berat badan kurang, kelebihan berat badan, dan kekurangan gizi. Untuk melihat nilai indeks tersebut, digunakan metode *z-score*.

2.3.2. *Knowledge Discovery in Databases (KDD)*

Penemuan pengetahuan dalam basis data atau *knowledge discovery in database* (KDD) adalah suatu proses yang kompleks yang mencakup pengambilan informasi yang lebih berharga, mudah dipahami dan baru dari dalam gudang data yang besar dan kompleks. Proses KDD dimulai dengan penetapan tujuan yang jelas dan diakhiri dengan evaluasi terhadap hasil yang ditemukan, yang melibatkan interpretasi hasil yang diperoleh dengan menghubungkannya dengan bidang ilmu

lainnya (Nofriansyah & Nurcahyo, 2015). Proses KDD dapat dilihat pada Gambar 2.1.



Gambar 2.1. Proses *knowledge discovery database* (KDD)

2.3.3. Proses *Knowledge Discovery in Database*

Proses *knowledge discovery in database* terdiri dari lima tahapan sebagai berikut:

1. Seleksi Data (*data understanding*)

Pemilihan data dilakukan sebagai tahap awal dalam proses *Knowledge Discovery in Database* (KDD). Data yang terpilih akan digunakan untuk proses klasifikasi multi-label *machine learning* dan disimpan dalam berkas terpisah dari basis data operasional.

2. Pembersihan Data (*data cleaning*)

Proses pembersihan data melibatkan mencari data yang hilang (*missing value*), membuang duplikasi data (*redundancy*), memeriksa data yang inkonsisten, dan koreksi kesalahan data, termasuk kesalahan cetak (tipografi). Selain itu, proses cek *outlier* dilakukan untuk mengecek data

yang sangat berbeda dengan data lainnya di dalam dataset, agar mendapatkan informasi lain yang relevan dan diperlukan untuk KDD.

3. Persiapan Data (*data preparation*)

Pada tahap ini, data yang belum memiliki entitas yang jelas ditransformasikan menjadi bentuk yang valid atau siap untuk proses klasifikasi *machine learning*.

4. Modeling (*machine learning*)

Proses ini melibatkan penerapan algoritma atau metode pencarian untuk mendapatkan pengetahuan dari *dataset* yang diteliti, khususnya dalam konteks penggunaan model *machine learning* untuk klasifikasi *multi-label* dengan *hyperparameter tuning*.

5. Evaluasi (*evaluation*)

Pada tahap ini, *output* yang mudah dimengerti dihasilkan berdasarkan proses klasifikasi untuk membentuk pola informasi. Metode klasifikasi digunakan adalah untuk memprediksi *dataset*, dimana prediksi melibatkan nilai hasil yang akan datang berdasarkan pola yang ditemukan (Triwidianti dkk., 2021).

2.3.4. *Machine Learning*

Machine Learning (Pembelajaran Mesin) adalah suatu teknik yang digunakan untuk menarik kesimpulan tentang data dengan pendekatan matematis. Inti dari *machine learning* adalah membuat model matematis yang merefleksikan pola-pola yang terdapat dalam data. Melalui *machine learning* memungkinkan

komputer atau program dapat memperoleh pengetahuan tanpa perlu diprogram secara eksplisit (Diantika & Firmanto, 2020).

Machine learning diibaratkan sebuah alat yang terdiri dari rumus matematika, yang kemudian digunakan tergantung pada konteks masalah yang dihadapi. Penggunaan *machine learning* memiliki minimal dua tujuan, yaitu meramalkan peristiwa masa depan yang belum diamati dan atau menemukan pengetahuan baru. Dalam *machine learning*, terdapat dua metode pembelajaran yang paling umum digunakan, pembelajaran yang diawasi (*Supervised learning*) dan pembelajaran yang tidak diawasi (*Unsupervised learning*).

2.3.5. Klasifikasi

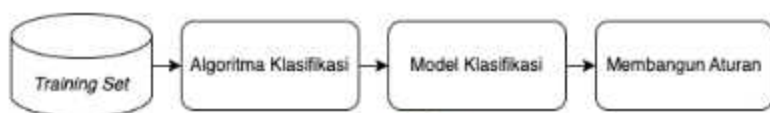
Algoritma *machine learning* dapat dikelompokkan menjadi tiga jenis, yaitu *supervised*, *unsupervised*, dan *semi-supervised* (Neelamegam & Ramaraj, 2013). Pada *supervised learning*, algoritma menggunakan data yang telah diberi label kelas. Dalam *unsupervised learning*, data tidak memiliki label kelas, dan algoritma digunakan untuk mengelompokkan data berdasarkan kesamaannya. Sedangkan pada *semi-supervised learning*, sebagian kecil data telah diberi label, sementara sebagian lainnya belum.

Proses klasifikasi dibagi menjadi dua tahap (Annasaheb & Verma, 2016), yaitu:

1. Tahap membangun model

Pada tahap ini, model klasifikasi dibangun berdasarkan data yang telah diberi label kelas. Data sampel yang digunakan untuk proses ini disebut sebagai

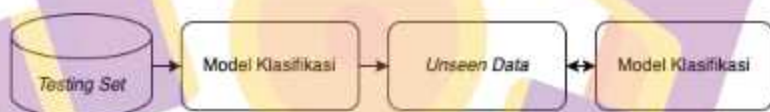
data pelatihan atau data pembelajaran (*training set*). Proses ini dikenal sebagai proses induksi yang ditunjukkan pada Gambar 2.2.



Gambar 2.2. Tahapan membangun model data latih (*data training*)

2. Tahap penggunaan model klasifikasi

Pada tahap ini, model klasifikasi diterapkan pada data yang belum diberi label kelasnya. Proses penerapan model klasifikasi untuk memprediksikan kelas label kelas dari data baru menggunakan data uji (*testing set*), proses ini disebut deduksi. Proses ini dapat dilihat pada Gambar 2.3.



Gambar 2.3. Tahapan menggunakan model data uji (*data test*)

2.3.6. Klasifikasi Multi-Label

Klasifikasi multi-label adalah bagian dari masalah klasifikasi teks, dimana setiap dokumen dapat termasuk dalam beberapa kelas sekaligus. Ini berbeda dengan klasifikasi *single-label* (biner), yang bertujuan untuk mengklasifikasikan dokumen ke dalam satu kelas saja (Isnaini dkk., 2019) (Pane dkk., 2018) (Khonsa Izzaty dkk., 2018). Dalam klasifikasi multi-label, setiap dokumen dalam data latih memiliki satu set label dengan tujuan memprediksi set label untuk dokumen yang kelasnya belum diketahui (Zhang dkk., 2009).

2.3.7. Naïve Bayes

Naïve bayes adalah algoritma klasifikasi yang sederhana namun efisien, yang terkenal karena kemampuannya dalam implementasi yang mudah, kinerja yang baik pada *dataset* besar, dan kemampuannya untuk menangani *multi-class classification*. Meskipun algoritma ini biasanya efektif pada data dengan banyak kelas dan dapat memberikan hasil yang memuaskan meskipun asumsi dasar independensi fitur yang sederhana, tidak selalu cocok untuk semua jenis data. Namun, *naïve bayes* tetap menjadi pilihan yang menarik terutama dalam konteks dimana asumsi dasarnya dapat diterapkan dan interpretasi model menjadi prioritas. Rumus dasar *naïve bayes* dapat dijelaskan menggunakan teorema *bayes*, seperti yang ditunjukkan pada persamaan 1.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

Dalam konteks klasifikasi, terdapat kelas target (A) dan fitur-fitur (B). *Naïve bayes* menghitung probabilitas kelas (C_k) untuk suatu data (X) berdasarkan fitur-fiturnya, sebagaimana dijelaskan dalam persamaan 2.

$$P(C_k|X) = \frac{P(X|C_k) \cdot P(C_k)}{P(X)} \quad (2)$$

Dimana ($C_k|X$) adalah probabilitas kelas C_k mengingat data X , $P(X|C_k)$ adalah *likelihood* dari data C_k , $P(C_k)$ adalah prior probabilitas dari kelas C_k , dan $P(X)$ adalah probabilitas margin dari data X . Asumsi dasar dari *naïve bayes* adalah bahwa fitur-fitur X bersifat independen jika diketahui kelasnya. Oleh karena itu, *likelihood* dapat dihitung sebagai perkalian dari probabilitas individu fitur, seperti yang ditunjukkan dalam persamaan 3.

$$P(C_k|X) = P(x_1|C_k) \cdot (x_{21}|C_k) \dots (x_{1n}|C_k) \quad (3)$$

2.3.8. Logistic Regression

Logistic Regression merupakan sebuah metode statistik yang bertujuan untuk menggambarkan hubungan antara variabel dependen yang bersifat kategorik dengan satu atau lebih variabel independen. Meskipun namanya mengandung istilah regresi, *logistic regression* sebenarnya merupakan algoritma klasifikasi yang digunakan untuk memprediksi probabilitas terjadinya suatu peristiwa dengan *output* berupa kelas diskrit, biasanya dalam bentuk biner seperti ya/tidak atau 0/1. Algoritma ini memanfaatkan fungsi logistik untuk memetakan variabel *input* ke dalam nilai probabilitas dalam rentang 0 hingga 1, yang kemudian diinterpretasikan sebagai prediksi kelas target. Rumus *logistic regression* untuk kasus biner ditunjukkan pada persamaan 4.

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (4)$$

Dimana $P(Y = 1|X)$ adalah probabilitas bahwa variabel target Y adalah 1 berdasarkan variabel X , e adalah konstanta logaritma $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ adalah koefisien yang harus diestimasi dari data latih, X_1, X_2, \dots, X_n adalah variabel *input*. Rumus ini menggunakan fungsi logistik (atau sigmoid) untuk mengubah nilai hasil dari regresi linear ke dalam rentang 0 hingga 1, yang diinterpretasikan sebagai probabilitas kelas positif. Estimasi koefisien (β) dilakukan dengan menggunakan teknik seperti metode maksimum *likelihood*.

2.3.9. Random Forest

Random forest adalah sebuah teknik yang dapat meningkatkan tingkat akurasi dengan membangun sejumlah pohon keputusan secara acak. Proses pembangunan pohon keputusan dimulai dengan simpul akar, kemudian berlanjut dengan simpul percabangan yang memiliki minimal dua *output*, dan diakhiri dengan simpul daun yang hanya memiliki satu *input*. Pada setiap simpul, keputusan diambil dengan cara mengukur tingkat ketidakhomogenan atribut menggunakan entropi dan menghitung *information gain* dan penggunaan atribut secara acak, *random forest* membangun pohon keputusan yang beragam, menghasilkan prediksi yang lebih kuat dan akurat. Untuk menghitung nilai entropi, digunakan persamaan 5 dan 6.

$$Entropy(Y) = - \sum_i p(c|Y) \log_2 p(c|Y) \quad (5)$$

Dimana Y adalah himpunan kasus, dan $p(c|Y)$ merupakan proporsi nilai Y terhadap kelas c .

$$Information\ Gain(Y, a) = Entropy(Y) - \sum_{v \in a} values \frac{Y_v}{Y_a} Entropy(Y_v) \quad (6)$$

Dimana $values \frac{Y_v}{Y_a}$ merupakan semua nilai yang mungkin dalam himpunan kasus a , Y_v adalah subkelas dari Y dengan kelas v yang berhubungan dengan kelas a . Selain itu, pemilihan atribut untuk menjadi simpul, baik sebagai akar maupun sebagai internal, didasarkan pada *information gain* tertinggi yang dimiliki oleh atribut yang tersedia. *Gain ratio* dihitung dengan membagi hasil dari perhitungan *information gain* dengan nilai *split information*. *Split information* yang digunakan

dapat ditemukan pada persamaan 7, sementara nilai *gain ratio* terlihat dalam persamaan 8.

$$\text{Split Information } (S, A) = \sum_i^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (7)$$

Dimana *Split Information* (S, A) adalah perkiraan nilai entropy dari variabel S yang memiliki kelas c . Sementara $\frac{|S_i|}{|S|}$ merepresentasikan probabilitas dari kelas i dalam atribut tersebut.

$$\text{Gain ratio } (S, A) = \frac{\text{Information } (S, A)}{\text{Split information } (S, A)} \quad (8)$$

2.3.10. Support Vector Classifier

Support vector classifier adalah salah satu algoritma *machine learning* yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini mencari garis atau permukaan pemisah optimal di antara kelas-kelas data dengan memanfaatkan vektor pendukung. Keunggulan *support vector classifier* terletak pada kemampuannya dalam menangani ruang berdimensi tinggi dan data yang tidak linier, serta dapat diterapkan pada masalah klasifikasi multi-kelas. Meskipun kompleks, *support vector classifier* sering digunakan dalam berbagai konteks seperti pengenalan pola, bioinformatika, dan analisis citra. Dalam konteks klasifikasi, *support vector classifier* mencari *hyperplane* terbaik yang dapat memisahkan dua kelas data. *Hyperplane* ini dipilih sedemikian rupa sehingga margin antara dua kelas maksimal. Formula dasar *hyperplane* dalam *support vector classifier* dapat ditunjukkan pada persamaan 9.

$$f(x) = \omega \cdot x + b \quad (9)$$

Dimana ω adalah vektor bobot yang tegak lurus terhadap *hyperlane*, x adalah vektor fitur *input*, dan b adalah bias atau pergeseran. *Support vector classifier* berusaha memaksimalkan margin, yaitu jarak antara *hyperlane* dan sampel data terdekat dari setiap kelas. Margin ini dihitung sebagai $\frac{2}{\|\omega\|}$ Dimana $\|\omega\|$ adalah norma Euclidean dari vektor bobot. Formulasi *support vector classifier* dengan kernel ditunjukkan pada persamaan 10.

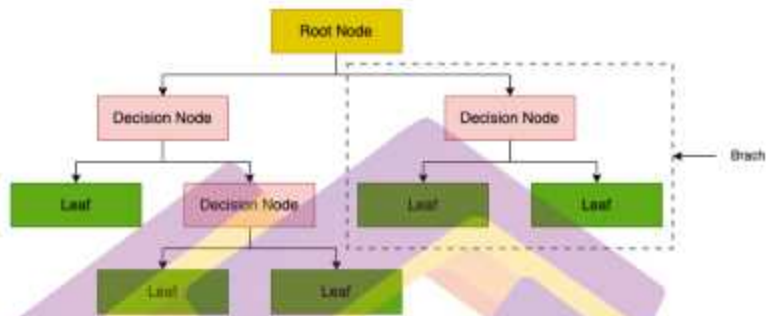
$$f(x) = \sum_{i=1}^N a_i y_i K(x, x_i) + b \quad (10)$$

Dimana K adalah fungsi kernel, a adalah bobot dari vektor dukungan, y adalah label kelas, dan b adalah bias. Untuk klasifikasi, prediksi kelas dilakukan dengan menghitung $f(x)$ dan menentukan kelas berdasarkan tanda dari $f(x)$ (positif atau negatif). Parameter *support vector classifier*, seperti ω , a , dan b , ditentukan melalui proses pelatihan yang melibatkan optimasi untuk mencapai margin maksimal dan mengatasi kesalahan klasifikasi.

2.3.11. Decision Tree

Decision tree adalah metode klasifikasi untuk prediksi yang mirip dengan struktur hirarki atau pohon (Adhinata & Junaidi, 2022). Menyusun urutan keputusan dan menghubungkan data ke pohon keputusan merupakan operasi dasar dari pohon keputusan. Salah satu manfaat utama dari menggunakan pohon keputusan adalah kemampuannya untuk menyederhanakan keputusan yang

kompleks. Proses ini memungkinkan pengambil keputusan untuk membuat konsep penyelesaian masalah yang lebih efisien (Noor dkk., 2019).



Gambar 2.4. *Decision tree*

Pohon keputusan terdiri dari simpul-simpul fundamental, yang dikenal sebagai *root node*, seperti yang dapat dilihat pada gambar 2.4. di atas. *Root node* menggambarkan seluruh kumpulan data dan berfungsi sebagai titik awal untuk proses pengambilan keputusan. *Node* keputusan/Internal merepresentasikan keputusan mengenai fitur *input*. *Leaf/Terminal node* adalah *node* yang tidak memiliki *node* anak dan mewakili label kelas atau nilai numerik. *Splitting* adalah proses pembagian sebuah *node* menjadi dua atau lebih sub-*node* berdasarkan kriteria pemisahan dan fitur yang dipilih. *Node* yang terbagi menjadi satu atau lebih *node* anak dikenal sebagai *parent node*. *Node* yang dihasilkan muncul ketika *parent node* dibagi disebut sebagai *child nodes* (Ali & Mashwani, 2023). Rumus algoritma *decision tree* dapat ditunjukkan pada persamaan 11.

$$H(T) = - \sum_{i=1}^c p(i). \log_2 p(i) \quad (11)$$

Dimana $H(T)$ adalah entropi dari pohon Keputusan T , tergantung pada jumlah kelas yang disimbolkan dengan c , dan proporsi sampel yang termasuk ke dalam kelas i dalam pohon keputusan T , yang dinyatakan sebagai $p(i)$.

2.3.12. *K-Nearest Neighbors*

K-nearest neighbors merupakan bagian dari algoritma *supervised machine learning* yang mudah diimplementasikan dan mampu menangani masalah yang kompleks. Namun, *k-nearest neighbors* sering disebut *lazy learning* karena memerlukan penyimpanan semua data pelatihan dan pengukuran kesamaannya terhadap setiap data baru yang akan diprediksi (Lonang, Yudhana, & Biddinika, 2023). Untuk mengklasifikasikan data baru yang belum memiliki kelas tertentu, algoritma *k-nearest neighbors* mengidentifikasi k (tetangga) titik data yang memiliki jarak paling dekat dengan data baru tersebut. Kelas prediksi untuk data baru ditentukan berdasarkan frekuensi kelas dari kumpulan data terdekat, yaitu k . Untuk mencegah jarak yang sama selama proses klasifikasi, nilai k biasanya diberikan nilai ganjil. Fungsi jarak seperti *Manhattan distance*, digunakan untuk menghitung jarak atau kedekatan antar tetangga (Pertiwi dkk., 2020). Rumus *Manhattan distance* dapat dilihat pada persamaan 12.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (12)$$

2.3.13. *Confusion Matrix*

Untuk mengevaluasi performa model klasifikasi, *confusion matrix* dapat digunakan dalam *machine learning* untuk menganalisis lebih mendalam kinerja

algoritma yang akan dibandingkan (Sahu dkk., 2019). *Confusion matrix* merupakan tabel yang umumnya digunakan untuk menggambarkan performa model klasifikasi pada sebuah set data uji yang nilai sebenarnya diketahui. Setelah *confusion matrix* berhasil diperoleh dari setiap algoritma yang diujikan, maka metrik seperti akurasi, presisi, *recall*, *f1-score*, serta kurva AUC-ROC dapat dihitung. Tabel *confusion matrix* dapat dilihat pada gambar 2.5.

Actual Class	0	True Negative (TN)	False Positive (FP)
	1	False Negative (FN)	True Positive (TP)
		0	1
		Predicted Class	

Gambar 2.5. Tabel *Confusion matrix*

Berdasarkan konvensi yang digunakan, TN atau *true negative* mencerminkan jumlah sampel negatif yang berhasil diklasifikasikan dengan benar, sementara TP atau *true positive* menunjukkan jumlah sampel positif yang berhasil diklasifikasikan secara tepat. Selain itu, FP atau *false positive* menggambarkan jumlah sampel negatif sebenarnya yang keliru diklasifikasikan sebagai positif, dan FN atau *false negative* mengindikasikan jumlah sampel positif yang salah diklasifikasikan sebagai negatif (Vanacore dkk., 2024) (Singh dkk., 2021).

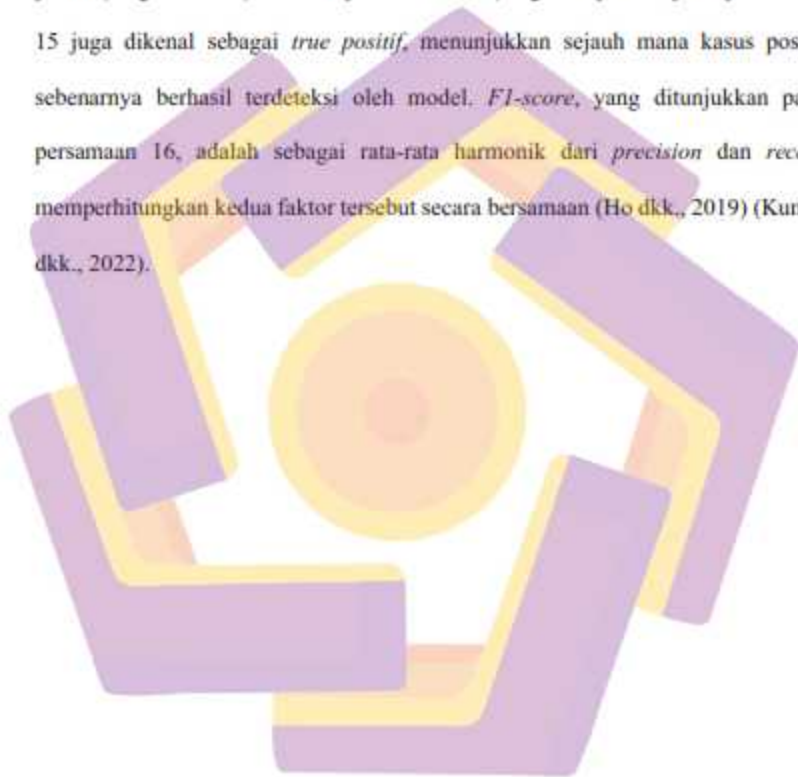
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (13)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (14)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (15)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (16)$$

Akurasi pada persamaan 13 mencerminkan persentase keseluruhan prediksi yang benar, sedangkan Presisi pada persamaan 14 mengukur sejauh mana prediksi positif yang sebenarnya adalah positif. *Recall*, yang ditunjukkan pada persamaan 15 juga dikenal sebagai *true positif*, menunjukkan sejauh mana kasus positif sebenarnya berhasil terdeteksi oleh model. *F1-score*, yang ditunjukkan pada persamaan 16, adalah sebagai rata-rata harmonik dari *precision* dan *recall*, memperhitungkan kedua faktor tersebut secara bersamaan (Hodgson, 2019) (Kumar, 2022).



BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

3.1.1. Jenis Penelitian

Penelitian ini merupakan sebuah eksperimen yang bertujuan untuk memberikan solusi terhadap permasalahan tertentu. Jenis penelitian ini termasuk dalam kategori kuantitatif karena menggunakan data berbentuk angka. Dalam penelitian ini, algoritma *machine learning* seperti *naïve bayes*, *logistic regression*, *random forest*, *svc*, *decision tree* dan *k-nearest neighbors* digunakan untuk melakukan klasifikasi multi-label dengan *hyperparameter tuning* terhadap status gizi seperti *stunting*, *wasting*, dan *underweight* pada balita.

3.1.2. Sifat Penelitian

Penelitian ini bertujuan untuk melakukan klasifikasi multi-label dengan *hyperparameter tuning* pada status gizi balita seperti *stunting*, *wasting*, dan *underweight*, dengan melakukan perbandingan algoritma *machine learning* yang mencakup *naïve bayes*, *logistic regression*, *random forest*, *svc*, *decision tree* dan *k-nearest neighbors*. Penelitian ini akan menggambarkan status gizi balita di Kabupaten Majalengka dengan menggunakan data yang diperoleh dari Dinas Kesehatan Kabupaten Majalengka. Dengan menggunakan pendekatan kausal komparatif, penelitian ini akan mengeksplorasi hubungan sebab-akibat antara penggunaan berbagai algoritma *machine learning* dan hasil klasifikasi status gizi

balita. Berdasarkan data yang ada, penelitian ini akan menghasilkan hasil klasifikasi deskriptif yang membagi balita menjadi kelas berdasarkan status gizinya.

3.1.3. Pendekatan Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif untuk membandingkan kinerja algoritma *machine learning* berdasarkan tingkat akurasi yang di dapat dari klasifikasi multi-label dengan *hyperparameter tuning* pada status gizi balita seperti *stunting*, *wasting*, dan *underweight*. Algoritma yang dievaluasi meliputi *naïve bayes*, *logistic regression*, *random forest*, *svc*, *decision tree* dan *k-nearest neighbors*. Pendekatan ini melibatkan pengumpulan data numerik terkait status gizi dan dilanjutkan dengan analisis statistik menggunakan masing-masing algoritma untuk menghasilkan model klasifikasi. Pendekatan kuantitatif bertujuan untuk memberikan informasi yang terukur dan objektif mengenai status gizi pada balita berdasarkan hasil analisis statistik yang dilakukan.

3.2. Metode Pengumpulan Data

Sebelum memperoleh data, permohonan penelitian diajukan ke Pascasarjana Universitas Amikom Yogyakarta dengan nomor 093/PASCASARJANA/AMIKOM/V/2023, pada tanggal 08 Mei 2023, yang ditujukan kepada Kepala Dinas Kesehatan Kabupaten Majalengka. Selain itu, permohonan serupa diajukan kepada Pimpinan Badan Kesatuan Bangsa dan Politik (BAKESBANGPOL) Kabupaten Majalengka dengan nomor 100/PASCASARJANA/AMIKOM/V/2023, pada tanggal 16 Mei 2023. Setelah itu,

diajukan ke kantor BAKESBANGPOL Kabupaten Majalengka untuk pembuatan surat pengantar dengan nomor surat: PP.05.01/201/Kesbang tanggal 22 Mei 2023.

Setelah surat permohonan penelitian dari universitas dan surat pengantar dari BAKESBANGPOL selesai, diterbitkan surat izin pendahuluan oleh Dinas Kesehatan Kabupaten Majalengka dengan nomor 071/3595/SDK tanggal 21 Juni 2023. Surat izin ini diberikan untuk mengizinkan penulis untuk melakukan pengambilan data guna penyusunan tesis.

Data yang digunakan diperoleh secara langsung dari Dinas Kesehatan Kabupaten Majalengka. Data tersebut adalah Rekapitulasi Hasil Bulan Penimbangan Balita (BPB) di Kabupaten Majalengka tahun 2022, mencakup kegiatan penimbangan berat badan dan pengukuran panjang atau tinggi badan pada balita usia 0-60 bulan. Kegiatan tersebut dilaksanakan secara serentak di Posyandu pada bulan Agustus 2022.

Dataset ini terdiri dari 3.504 data dengan 10 atribut. Dari jumlah tersebut, 7 atribut merupakan atribut fitur, sedangkan 3 atribut lainnya merupakan label. Atribut fitur mencakup jenis kelamin, berat badan ketika lahir, tinggi badan ketika lahir, usia, berat badan, tinggi badan, dan lingkaran lengan atas. Sementara itu, atribut label ini terdiri dari tinggi badan per umur, berat badan per umur, dan berat badan per tinggi badan. Atribut label ini mencerminkan status gizi balita, dimana *stunting* diklasifikasikan dengan kelas normal, pendek, dan sangat pendek. *Wasting* diklasifikasikan dengan kelas berat badan normal, kurang, sangat kurang, dan risiko lebih. Sedangkan *underweight* diklasifikasikan dengan kelas gizi baik, gizi kurang, risiko gizi lebih, gizi lebih, obesitas, dan gizi buruk.

3.3. Metode Analisis Data

Metode analisis data dalam penelitian ini melibatkan beberapa tahapan utama yang dilakukan secara sistematis untuk memastikan kualitas dan akurasi model klasifikasi yang dibangun. Tahapan pertama adalah pemahaman data (*data understanding*), dimana struktur dan isi *dataset* yang digunakan analisis untuk memastikan kesesuaiannya. *Dataset* yang digunakan memuat atribut fitur seperti jenis kelamin, berat badan ketika lahir, tinggi badan ketika lahir, usia, berat badan, tinggi badan, dan lingkaran lengan atas, serta atribut label yang terdiri dari tinggi badan per umur (tb/u), berat badan per umur (bb/u), dan berat badan per tinggi badan (bb/tb).

Langkah berikutnya adalah pembersihan data (*data cleaning*), yang mencakup pengecekan dan penanganan *missing values*, penghapusan data duplikat, identifikasi dan penggantian nilai 0 yang tidak sesuai, serta deteksi dan penanganan *outlier*. Tahapan ini penting untuk memastikan integritas data yang akan digunakan dalam analisis selanjutnya. Analisis data eksploratif (*exploratory data analysis*, *EDA*) kemudian dilakukan untuk memahami distribusi data dan hubungan data antar atribut. Berbagai visualisasi data seperti histogram digunakan untuk menampilkan distribusi usia, berat badan, dan tinggi badan balita. Analisis dilakukan untuk melihat distribusi label *stunting*, *wasting*, dan *underweight* berdasarkan jenis kelamin. Selain itu, *scatter plot* digunakan untuk menganalisis hubungan antara berat dan tinggi balita dan *correlation matrix*.

Tahap persiapan data (*data preparation*) meliputi beberapa langkah penting, termasuk *label encoding* untuk mengubah label kategorik menjadi numerik, dan

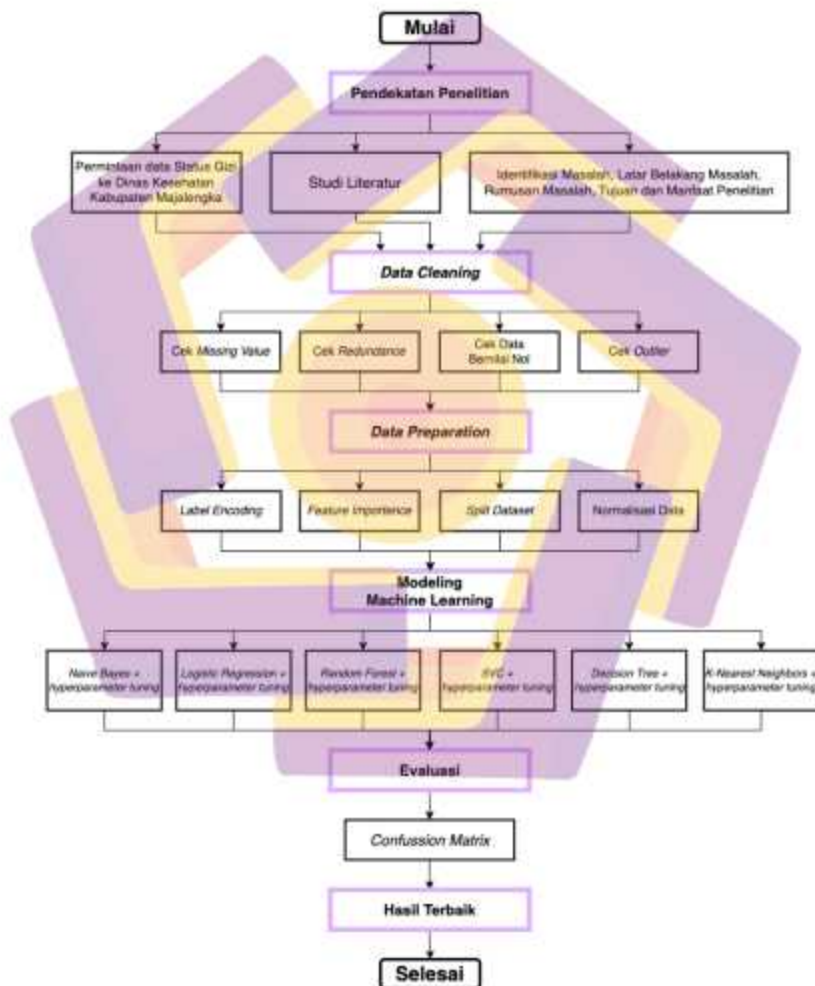
penentuan pentingnya fitur (*feature importance*). Data kemudian dipisahkan menjadi data latih dan data uji (*train-test split*), dan normalisasi data dilakukan untuk memastikan bahwa setiap fitur memiliki skala yang sama, yang sangat penting dalam meningkatkan kinerja model *machine learning*.

Pada tahap pemodelan (*modelling*), beberapa algoritma *machine learning* digunakan untuk membangun model klasifikasi multi-label. Algoritma yang digunakan meliputi *naïve bayes*, *logistic regression*, *random forest*, *svc*, *decision tree* dan *k-nearest neighbors*. Untuk setiap model, dilakukan *hyperparameter tuning* menggunakan *GridSearchCV* guna memperoleh parameter terbaik yang dapat meningkatkan performa model.

Evaluasi model dilakukan dengan menggunakan beberapa metrik, termasuk akurasi yang mengukur tingkat akurasi model, *classification report* yang memberikan gambaran tentang *precision*, *recall*, dan *f1-score*, serta *confusion matrix* yang menunjukkan jumlah prediksi benar dan salah untuk setiap kelas. Evaluasi ini penting untuk memahami kinerja model secara mendalam. Tahapan akhir adalah pengujian data aktual (*testing* dan aktual), dimana hasil klasifikasi diuji dengan nilai sebenarnya untuk setiap model dan label. Perbandingan ini melibatkan identifikasi jumlah nilai yang berbeda dan indeks dimana nilai berbeda. Langkah ini penting untuk memastikan bahwa model yang dibangun tidak hanya memiliki performa yang baik pada data pelatihan, tetapi juga mampu memberikan hasil yang akurat pada data baru yang pernah dilihat oleh model.

3.4. Alur Penelitian

Penelitian ini membandingkan algoritma *machine learning* untuk klasifikasi multi-label dengan *hyperparameter tuning*. Adapun alur penelitiannya dapat dilihat pada Gambar 3.1.



Gambar 3.1. Alur Penelitian

Gambar 3.1 menjelaskan bahwa penelitian ini dilakukan dalam enam tahapan yaitu:

Alur penelitian dimulai dengan pendekatan penelitian yang mencakup permintaan data, studi literatur untuk memahami konteks dan identifikasi masalah, serta penentuan latar belakang masalah, perumusan masalah, tujuan penelitian, dan manfaat penelitian. Langkah ini memastikan bahwa penelitian memiliki dasar yang kuat dan relevan.

Selanjutnya, penelitian berlanjut ke tahap persiapan data yang terdiri dari sub-langkah penting. Tahap pertama adalah pembersihan data (*data cleaning* atau *preprocessing*), yang meliputi pengecekan nilai-nilai yang hilang (*missing value*), pengecekan data duplikat, pengecekan data yang memiliki nilai nol, dan identifikasi serta penanganan *outlier*. Langkah-langkah ini penting untuk memastikan integritas dan kualitas data yang akan digunakan analisis lebih lanjut.

Tahap berikutnya adalah analisis eksploratif data (EDA – *Exploratory Data Analysis*) dan visualisasi. Ini termasuk pembuatan plot histogram untuk fitur-fitur utama seperti usia, berat badan, dan tinggi badan balita, analisis *stunting*, *wasting*, dan *underweight* berdasarkan jenis kelamin, pembuatan *scatter plot* antara berat dan tinggi balita, serta pembuatan *correlation matrix*. EDA memberikan gambaran awal tentang distribusi data dan hubungan antar variabel.

Setelah EDA, tahap persiapan data dilanjutkan dengan beberapa langkah penting yaitu *label encoding* untuk menangani data kategorikal, analisis *feature importance* untuk menentukan fitur-fitur yang paling berpengaruh, pembagian data

menjadi data latih dan data uji (*train-test split*), serta normalisasi data untuk memastikan bahwa semua fitur berada dalam skala yang sama.

Tahap modeling adalah inti dari penelitian ini, dimana berbagai algoritma *machine learning* digunakan untuk membangun model klasifikasi multi-label. Algoritma yang digunakan meliputi *naïve bayes*, *logistic regression*, *random forest*, *svc*, *decision tree*, dan *k-nearest neighbors*. Setiap model mengalami proses *hyperparameter tuning* untuk mengoptimalkan kinerjanya. Hasil yang dicari dari tahap ini adalah tingkat akurasi, parameter terbaik, klasifikasi model dan *classification report*.

Tahap evaluasi dilakukan menggunakan *confusion matrix* untuk memahami kinerja model lebih lanjut. Hasil evaluasi mencakup seperti akurasi (*accuracy*), presisi (*precision*), *recall*, dan *f1-score* untuk masing-masing model yang dibangun menggunakan algoritma *naïve bayes*, *logistic regression*, *random forest*, *svc*, *decision tree* dan *k-nearest neighbors*. Langkah terakhir adalah menguji hasil klasifikasi model dengan data aktual dari setiap model, guna membandingkan prediksi model dengan data aktual dan memahami kesalahan klasifikasi yang terjadi. Penelitian ini diakhiri dengan menyimpulkan hasil terbaik dari seluruh proses yang telah dilakukan.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Data Understanding

Penelitian ini menggunakan *dataset* status gizi balita yang diperoleh dari Dinas Kesehatan Kabupaten Majalengka, yang terdiri dari 3.504 data. *Dataset* ini terdiri dari 10 atribut mencakup 7 fitur yaitu jenis kelamin, berat badan ketika lahir, panjang atau tinggi badan ketika lahir, usia, berat badan, tinggi badan, dan lingkar lengan atas, serta 3 label yaitu tinggi badan per umur, berat badan per umur, dan berat badan per tinggi badan. *Dataset* status gizi dapat dilihat pada gambar 4.1.

	jk	bb_lahir	tb_lahir	usia	berat	tinggi	lila	tb/u	bb/u	bb/tb
0	P	3.4	50	30	12.4	86.8	16.0	Normal	Berat Badan Normal	Gizi Baik
1	L	2.3	45	38	12.8	99.5	15.0	Normal	Kurang	Gizi Kurang
2	P	2.7	47	55	14.0	103.8	15.5	Normal	Berat Badan Normal	Gizi Baik
3	L	2.8	48	54	12.2	96	14.0	Pendek	Kurang	Gizi Kurang
4	L	2.8	48	53	14.2	98.3	16.0	Normal	Berat Badan Normal	Gizi Baik
...
3499	L	3.3	50	36	13.7	82	17.0	Normal	Berat Badan Normal	Gizi Baik
3500	P	3.6	53	8	7.3	69.6	15.0	Normal	Berat Badan Normal	Gizi Baik
3501	P	2.9	48	7	6.9	69.1	14.0	Normal	Berat Badan Normal	Gizi Baik
3502	L	3	49	0	3.0	49	NaN	Normal	Berat Badan Normal	Gizi Baik
3503	L	2.8	48	0	2.8	48	NaN	Normal	Berat Badan Normal	Gizi Baik

3504 rows x 10 columns

Gambar 4.1. *Dataset* status gizi

Atribut dari ketiga label tersebut, *tb/u* (tinggi badan per umur) menunjukkan status gizi *stunting* yang terdiri dari kelas normal, pendek, dan sangat pendek. Label *bb/u* (berat badan per umur) menunjukkan status gizi *wasting* yang terdiri dari kelas berat badan normal, kurang, sangat kurang, dan risiko lebih. Sementara itu, label

bb/tb (berat badan per tinggi badan) menunjukkan status gizi *underweight* yang terdiri dari kelas gizi baik, gizi kurang, risiko gizi lebih, gizi lebih, obesitas, dan gizi buruk. Penjelasan lebih lanjut mengenai atribut-atribut dalam *dataset* terdapat pada tabel 4.1.

Tabel 4.1. Deskripsi *dataset*

No.	Atribut	Deskripsi
1.	jk	Jenis Kelamin (L : Laki-laki, P : Perempuan)
2.	bb_lahir	Berat badan ketika lahir (kg)
3.	tb_lahir	Panjang atau tinggi badan ketika lahir (cm)
4.	usia	Usia balita (0-60 bulan)
5.	berat	Berat badan balita ketika ditimbang (kg)
6.	tinggi	Tinggi badan balita ketika diukur (cm)
7.	lila	Lingkar lengan atas balita (cm)
8.	tb/u	Tinggi badan per umur (status gizi <i>stunting</i>), dengan kelas: - Normal - Pendek - Sangat pendek
9.	bb/u	Berat badan per umur (status gizi <i>wasting</i>), dengan kelas: - Berat badan normal - Kurang - Sangat kurang - Risiko lebih
10.	bb/tb	Berat badan per tinggi badan (status gizi <i>underweight</i>), dengan kelas: - Gizi baik - Gizi kurang - Risiko gizi lebih - Gizi lebih - Obesitas - Gizi buruk

Dataset yang digunakan dalam penelitian ini terdiri dari tiga tipe data, yaitu *object*, *float*, dan *integer*. Atribut-atribut seperti *jk*, *bb_lahir*, *tb_lahir*, *tinggi*, *tb/u*, *bb/u*, dan *bb/tb*, memiliki tipe data *object*. Atribut *usia* memiliki tipe data *integer*, sedangkan atribut berat badan balita ketika ditimbang dan *lila* memiliki tipe data *float*. Tipe data dapat dilihat pada gambar 4.2.

```

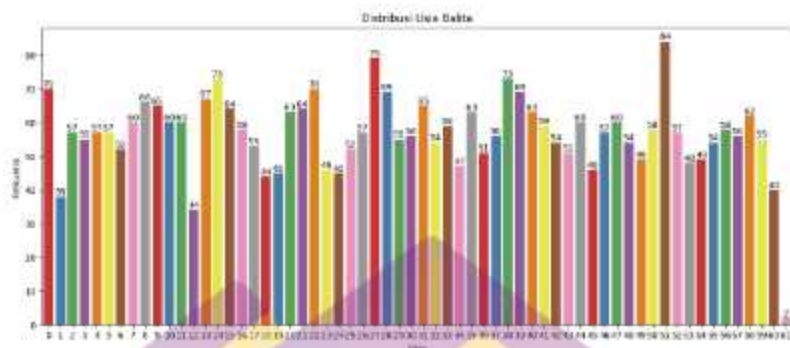
jk      object
bb_lahir object
tb_lahir object
usia    int64
berat   float64
tinggi  object
lla     float64
tb/u    object
bb/u    object
bb/tb   object

dtype: object

```

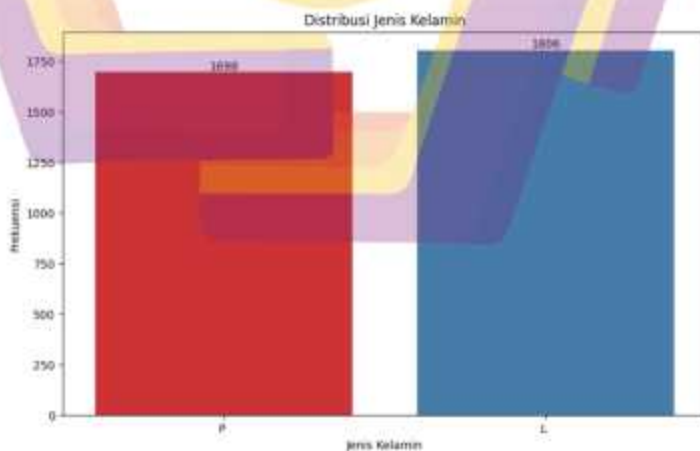
Gambar 4.2. Tipe data status gizi

Grafik distribusi usia balita menunjukkan gambaran visual mengenai distribusi usia balita dalam suatu populasi. Setiap batang pada grafik mewakili satu kelompok usia (dalam bulan), dan tinggi batang menunjukkan frekuensi atau jumlah balita yang termasuk dalam kelompok usia tersebut. Rentang usia balita dalam *dataset* mulai dari usia 0 sampai 61 bulan. Secara keseluruhan, terdapat fluktuasi signifikan dalam jumlah balita di berbagai usia. Puncak tertinggi terlihat pada usia 51 bulan dengan 84 balita, sementara beberapa usia lainnya menunjukkan jumlah yang relatif stabil atau mengalami penurunan drastis. Distribusi ini memberikan gambaran awal mengenai karakteristik populasi balita dalam *dataset*, yang dapat berguna untuk analisis lebih lanjut mengenai status gizi berdasarkan kelompok usia tertentu. Distribusi usia balita dapat dilihat pada gambar 4.3.



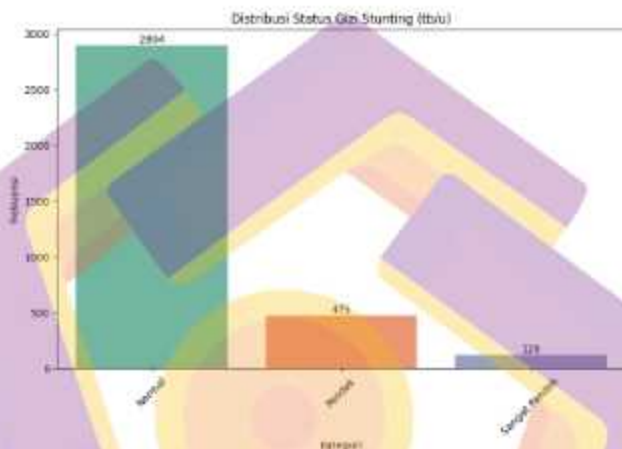
Gambar 4.3. Distribusi usia balita

Grafik distribusi jenis kelamin menunjukkan terdapat 1.698 balita berjenis kelamin perempuan dan 1.806 balita berjenis kelamin laki-laki. Jumlah balita dengan jenis kelamin laki-laki sedikit lebih banyak dibandingkan dengan perempuan. Perbedaan jumlah ini memberikan wawasan mengenai distribusi jenis kelamin dalam populasi balita yang diteliti. Distribusi jenis kelamin dapat dilihat pada gambar 4.4.



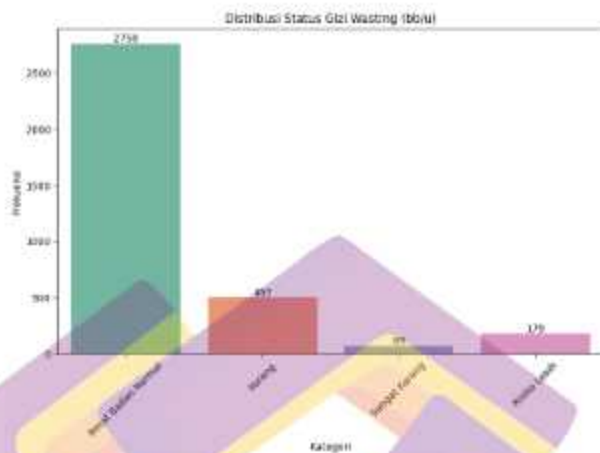
Gambar 4.4. Distribusi jenis kelamin balita

Grafik distribusi status gizi balita terkait *stunting*, data menunjukkan bahwa terdapat 2.894 balita dengan tinggi badan tergolong normal, 475 balita dengan tinggi badan kelas pendek, dan 120 balita yang tergolong sangat pendek. Distribusi status gizi *stunting* dapat dilihat pada gambar 4.5.



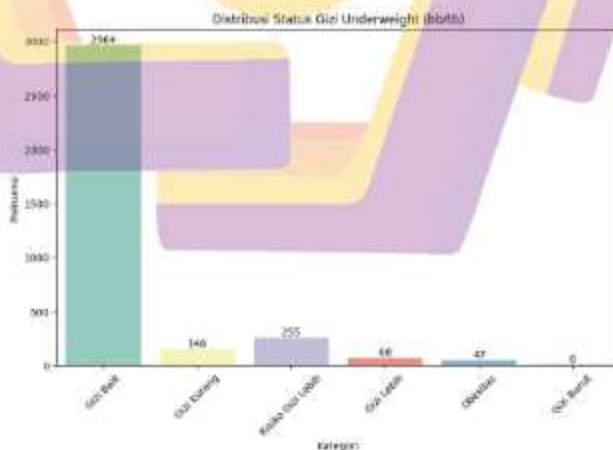
Gambar 4.5. Distribusi status gizi *stunting*

Grafik distribusi status gizi balita terkait *wasting*, data menunjukkan bahwa terdapat 2.759 balita dengan berat badan yang tergolong normal, 497 balita mengalami berat badan kurang, 69 balita dengan berat badan sangat kurang, dan 179 balita berada pada kelas risiko lebih. Distribusi status gizi *wasting* dapat dilihat pada gambar 4.6.



Gambar 4.6. Distribusi status gizi *wasting*.

Distribusi status gizi balita *underweight*, data menunjukkan bahwa terdapat 2.964 balita dengan kondisi gizi yang baik, 148 balita gizi kurang, 255 balita risiko gizi lebih, 68 balita gizi lebih, 47 balita tergolong obesitas, dan 6 balita mengalami gizi buruk. Distribusi status gizi *underweight* dapat dilihat pada gambar 4.7.



Gambar 4.7. Distribusi status gizi *underweight*

4.2. Data Cleaning

Setelah memahami karakteristik data, tahapan selanjutnya adalah melakukan *preprocessing* melalui pembersihan data untuk memastikan kualitas data yang digunakan dalam perbandingan klasifikasi menggunakan algoritma *machine learning*. Proses pembersihan data mencakup beberapa langkah sebagai berikut:

1. Cek data terhadap nilai yang hilang (*missing value*)

Pada tahap ini, dilakukan pengecekan terhadap nilai yang hilang pada setiap atribut *dataset* untuk memastikan tidak ada data yang hilang. Ditemukan bahwa terdapat nilai yang hilang pada atribut *bb_lahir*, *tb_lahir*, *lila*, *tb/u*, dan *bb/tb*. Data yang hilang dapat mempengaruhi hasil analisis, mengingat pentingnya informasi dari setiap atribut dalam menentukan status gizi balita, baris data yang memiliki nilai *missing value* pada atribut-atribut ini akan dihapus untuk menjaga integritas data. Dengan cara ini, kualitas *dataset* dapat dipertahankan agar dapat memberikan hasil yang valid. Nilai yang hilang pada setiap atribut dapat dilihat pada gambar 4.8.

```
Jumlah missing value:
jk          0
bb_lahir    2
tb_lahir    2
usia        0
berat       0
tinggi      0
lila        69
tb/u        15
bb/u        0
bb/tb       16
dtype: int64
```

Gambar 4.8. Nilai yang hilang pada setiap atribut

2. Cek data duplikat

Data duplikat dapat menyebabkan bias dalam analisis. Oleh karena itu, dilakukan pengecekan terhadap kemungkinan adanya data duplikat untuk mencegah redundansi yang tidak diinginkan (*redundancy*) serta memastikan konsistensi data pada *dataset*. Hasilnya, tidak ditemukan data duplikat dalam *dataset*, sehingga memastikan bahwa setiap entri data bersifat unik.

3. Cek data bernilai Nol (0)

Nilai nol (0) pada beberapa atribut dapat memiliki interpretasi yang berbeda. Pada *dataset* yang tersedia, ditemukan bahwa terdapat data yang bernilai nol pada atribut usia dan lila. Nilai nol pada atribut-atribut ini, dapat diindikasikan bahwa pendataan usia dan pengukuran lingkaran lengan atas tidak dilakukan. Keberadaan nilai tersebut harus rasional, dan oleh karena itu baris data yang memiliki nilai nol pada atribut tersebut akan dihapus. Data yang bernilai nol dapat dilihat pada gambar 4.9.

```
Data bernilai 0 pada atribut:
jk: 0
bb_lahir: 0
tb_lahir: 0
usia: 6
berat: 0
tinggi: 0
lila: 311
```

Gambar 4.9. Data bernilai nol (0)

4. *Cleaning data*

Proses *cleaning data* dilakukan untuk menghapus data nilai yang hilang dan data bernilai nol, jumlah *dataset* yang semula berjumlah 3.504 data berkurang menjadi 3.108 data. Hasil *cleaning data* untuk nilai yang hilang (*missing value*)

dapat dilihat pada gambar 4.10., sedangkan data dengan nilai nol dapat dilihat pada gambar 4.11:

```
Jumlah missing value:
jk          0
bb_lahir   0
tb_lahir   0
usia       0
berat      0
tinggi     0
lila       0
tb/u       0
bb/u       0
bb/tb      0
dtype: int64
```

Gambar 4.10. Nilai yang hilang setelah *cleaning data*

```
Data bernilai 0 pada kolom:
jk: 0
bb_lahir: 0
tb_lahir: 0
usia: 0
berat: 0
tinggi: 0
lila: 0
```

Gambar 4.11. Data bernilai nol setelah *cleaning data*

5. Deteksi *Outlier*

Sebelum deteksi *outlier*, perubahan tipe data pada atribut *bb_lahir*, *tb_lahir*, dan *tinggi* dilakukan dari tipe data *object* menjadi *float*, hal ini dilakukan untuk melakukan operasi matematis atau statistik pada atribut tersebut. Dengan konversi ini, analisis data seperti menghitung rata-rata, standar deviasi, atau visualisasi data dapat dilakukan dengan benar. Sehingga memungkinkan pemahaman yang lebih baik tentang *dataset* sebelum melakukan analisis lebih lanjut atau membangun model klasifikasi. Berikut tipe data setelah dilakukan konversi dapat dilihat pada gambar 4.12.

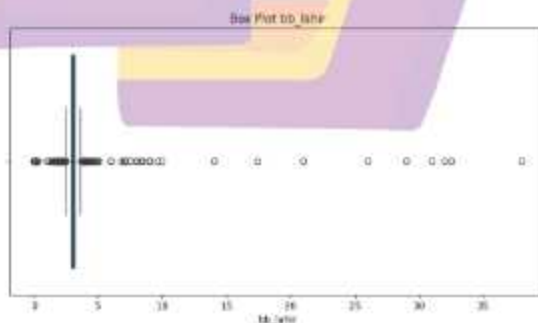
```

0
jk      object
bb_lahir float64
tb_lahir float64
usia    int64
berat   float64
tinggi  float64
lila    float64
tb/u    object
bb/u    object
bb/tb   object
dtype: object

```

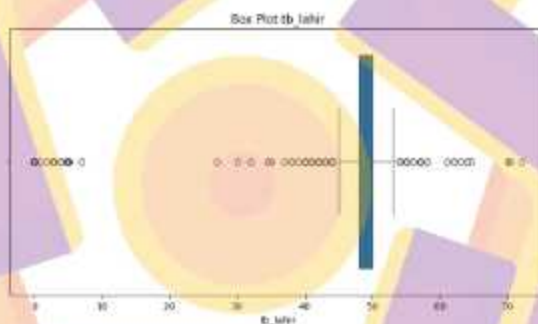
Gambar 4.12. Konversi tipe data *object* ke *float64*

Box plot pada gambar 4.13, menyajikan distribusi data dari atribut *bb_lahir*. Sebagian besar data *bb_lahir* berada pada berat lebih dari 2 kg sampai kurang dari 5 kg, ini menunjukkan bahwa berat badan bayi lahir umumnya berada pada rentang tersebut. Terdapat beberapa *outlier* pada berat yang lebih tinggi di atas 15 kg, ini menunjukkan adanya beberapa bayi yang memiliki berat badan lahir yang jauh di atas rata-rata atau bayi yang sudah berumur beberapa bulan tetapi baru dilakukan pengukuran untuk mengisi data berat badan lahir.



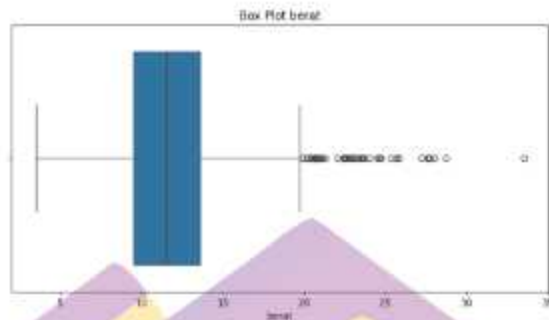
Gambar 4.13. *Box plot* berat badan lahir

Selanjutnya box plot pada gambar 4.14. menyajikan distribusi data dari atribut `tb_lahir`. Sebagian besar data `tb_lahir` berada pada panjang badan kurang lebih 45 cm sampai kurang lebih 53 cm, ini menunjukkan bahwa panjang badan bayi lahir umumnya berada pada rentang tersebut. Terdapat *outlier* pada panjang badan bayi lahir di bawah 30 cm dan beberapa di atas 60 cm, ini menunjukkan ada beberapa bayi yang memiliki panjang badan lahir yang jauh di atas atau di bawah rata-rata, *outlier* ini bisa terjadi karena kesalahan dalam pengumpulan data, misalnya kesalahan pencatatan atau kesalahan alat ukur.



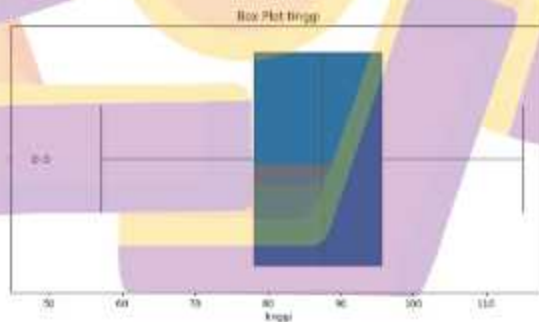
Gambar 4.14. *Box plot* panjang atau tinggi badan lahir

Selanjutnya box plot pada gambar 4.15. menyajikan distribusi data dari atribut berat. Sebagian besar data berat badan bayi berada pada berat 3 kg sampai 20 kg, ini menunjukkan bahwa berat badan bayi pada saat pengukuran umumnya berada pada rentang tersebut. Terdapat beberapa bayi berada pada berat di atas 25 kg, yang bisa dikategorikan sebagai *outlier*.



Gambar 4.15. *Box plot* berat badan bayi

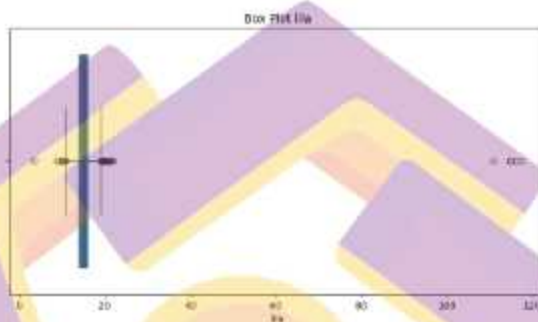
Selanjutnya box plot pada gambar 4.16. menyajikan distribusi data dari atribut tinggi. Sebagian besar data tinggi badan bayi berada pada tinggi 58 cm sampai lebih dari 110 cm, ini menunjukkan bahwa tinggi badan bayi pada saat pengukuran umumnya berada pada rentang tersebut. Terdapat beberapa bayi berada pada tinggi kurang dari 50 cm, yang bisa dikategorikan sebagai *outlier*.



Gambar 4.16. *Box plot* tinggi badan bayi

Selanjutnya box plot pada gambar 4.17. menyajikan distribusi data dari atribut lila. Sebagian besar data lingkaran lengan atas berada pada ukuran kurang lebih 10 cm sampai 20 cm, ini menunjukkan bahwa lingkaran lengan atas bayi pada

saat pengukuran umumnya berada pada rentang tersebut. Terdapat *outlier* beberapa bayi berada pada ukuran lebih dari 100 cm, *outlier* ini bisa terjadi karena kesalahan dalam pengumpulan data, misalnya kesalahan pencatatan atau kesalahan alat ukur.



Gambar 4.17. *Box plot* lingkaran lengan atas

Setelah dilakukan pengecekan *outlier* yang divisualisasikan dengan box plot, data yang nilainya sangat jauh berbeda dari data lainnya akan dihapus, karena dapat mempengaruhi hasil analisis. Ini bisa terjadi karena kesalahan pengukuran, kesalahan input data, atau memang kejadian langka. Sehingga jumlah *dataset* saat ini menjadi 2.968 data.

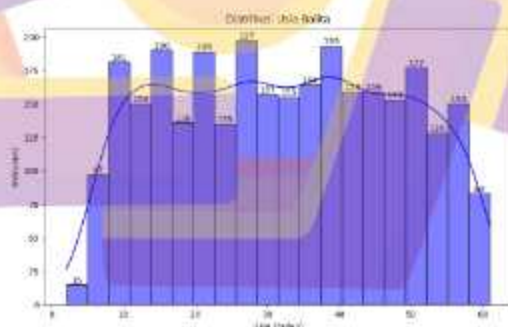
4.3. *Exploratory Data Analysis*

Analisis data eksploratif (EDA) adalah proses memahami *dataset* melalui visualisasi data. EDA bertujuan untuk menemukan pola, keterkaitan, anomali, dan mendapatkan wawasan penting dari data. Dalam tahap ini, fokus analisis terletak pada distribusi usia, berat badan, dan tinggi badan balita. Ini juga mencakup distribusi *stunting*, *wasting*, dan *underweight* per jenis kelamin. Selain itu, *scatter*

plot akan digunakan untuk memvisualisasikan hubungan antara berat badan dan tinggi badan balita, yang akan membantu memahami korelasi antara kedua variabel tersebut. Selain itu, *correlation matrix* akan digunakan menunjukkan seberapa erat hubungan antara berbagai atribut yang ada dalam dataset. diharapkan visualisasi ini akan meningkatkan pemahaman tentang data dan hubungan antar variabel, yang akan mendukung analisis lebih lanjut pada tahap selanjutnya. Berikut adalah visualisasi yang dilakukan pada tahap EDA:

1. Distribusi usia balita

Distribusi usia balita divisualisasi dalam bentuk histogram, yang menunjukkan penyebaran usia balita dari 0 bulan sampai 60 bulan. Distribusi usia balita tampak relatif merata di seluruh rentang usia tersebut, dengan jumlah terendah kurang lebih 15 balita dan tertinggi kurang lebih 197 balita. Grafik distribusi usia balita dapat dilihat pada gambar 4.18.

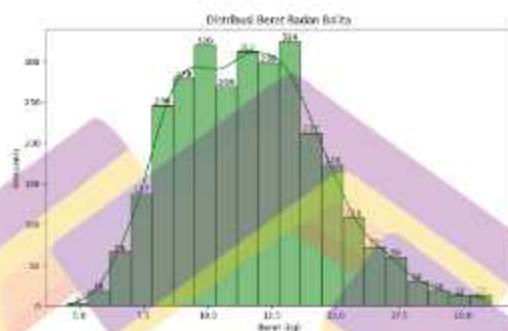


Gambar 4.18. Distribusi usia balita

2. Distribusi berat badan balita

Distribusi berat badan balita divisualisasi dalam bentuk histogram, yang menunjukkan penyebaran berat badan yang kurang lebih berkisar 5 kg sampai

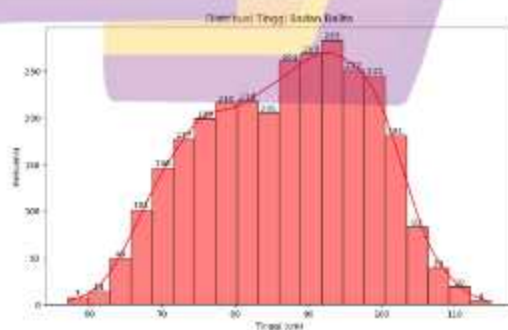
20 kg dengan berat badan yang bervariasi. Sebagian besar balita berada dalam rentang berat badan terbanyak pada rentang berat 8 kg sampai 15 kg. Grafik distribusi berat badan balita dapat dilihat pada gambar 4.19.



Gambar 4.19. Distribusi berat badan balita

3. Distribusi tinggi badan balita

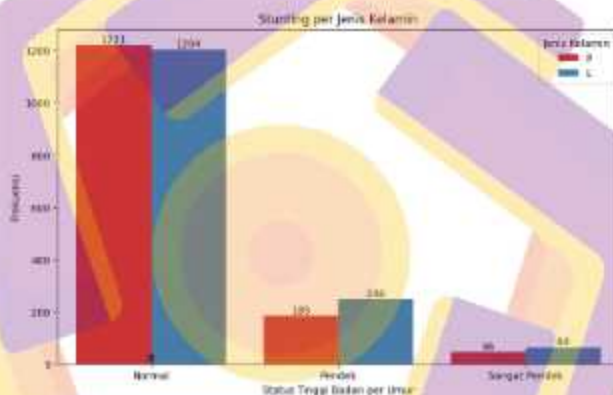
Distribusi tinggi badan balita divisualisasi dalam bentuk histogram, yang menunjukkan sebaran tinggi badan balita yang kurang lebih berkisar 60 cm sampai 110 cm. Sebagian besar balita memiliki tinggi badan dalam rentang tinggi kurang lebih 90 cm sampai 100 cm. Grafik distribusi tinggi badan dapat dilihat pada gambar 4.20.



Gambar 4.20. Distribusi tinggi badan balita

4. Distribusi *stunting* berdasarkan jenis kelamin

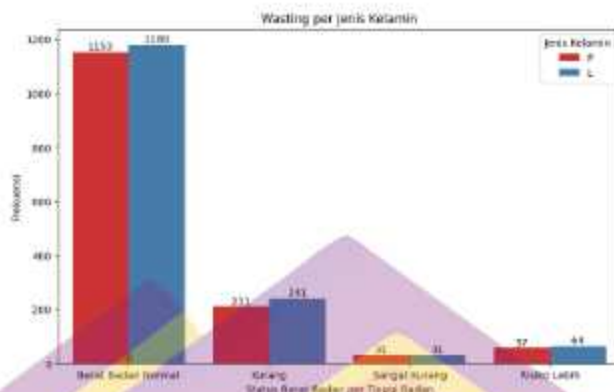
Distribusi status gizi tinggi badan per umur (tb/u) atau *stunting* berdasarkan jenis kelamin divisualisasikan menggunakan histogram. Hasilnya menunjukkan bahwa terdapat kecenderungan jumlah balita berjenis kelamin laki-laki yang mengalami *stunting*, baik dalam kelas pendek maupun sangat pendek, lebih tinggi dibandingkan dengan balita perempuan. Grafik distribusi *stunting* berdasarkan jenis kelamin dapat dilihat pada gambar 4.21.



Gambar 4.21. Distribusi *stunting* berdasarkan jenis kelamin

5. Distribusi *wasting* berdasarkan jenis kelamin

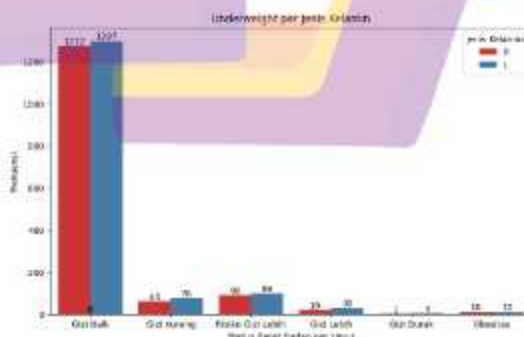
Distribusi status gizi berat badan per umur (bb/u) atau *wasting* berdasarkan jenis kelamin divisualisasikan menggunakan histogram. Hasilnya menunjukkan bahwa terdapat kecenderungan jumlah balita berjenis kelamin laki-laki yang mengalami *wasting*, baik dalam kelas kurang maupun risiko lebih, lebih tinggi dibandingkan balita perempuan. Grafik distribusi *wasting* berdasarkan jenis kelamin dapat dilihat pada gambar 4.22.



Gambar 4.22. Distribusi *wasting* berdasarkan jenis kelamin

6. Distribusi *underweight* berdasarkan jenis kelamin

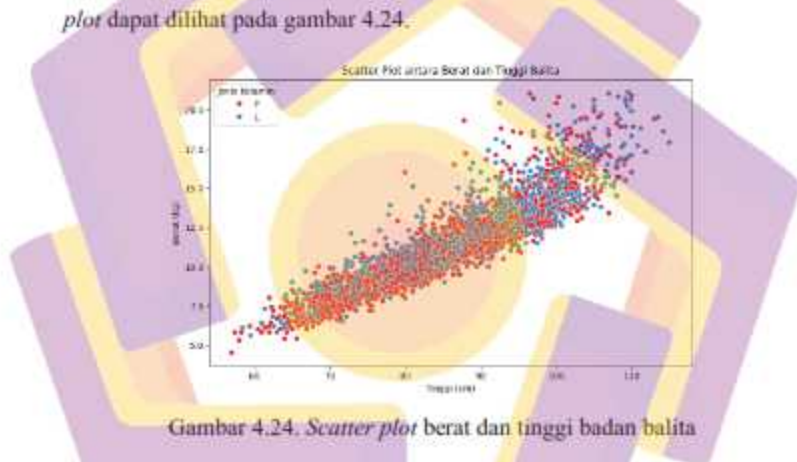
Distribusi status gizi berat badan per tinggi badan (bb/tb) atau *underweight* berdasarkan jenis kelamin divisualisasikan menggunakan histogram. Hasilnya menunjukkan bahwa terdapat kecenderungan balita berjenis kelamin laki-laki yang mengalami *underweight*, terutama pada kelas gizi kurang dan risiko gizi lebih, lebih tinggi dibandingkan dengan balita perempuan. Grafik distribusi *underweight* berdasarkan jenis kelamin dapat dilihat pada gambar 4.23.



Gambar 4.23. Distribusi *underweight* berdasarkan jenis kelamin

7. Scatter Plot berdasarkan berat dan tinggi balita

Scatter plot menggambarkan hubungan antara berat badan dan tinggi badan pada kelompok balita, dengan membedakan data berdasarkan jenis kelamin. Terdapat kecenderungan hubungan positif antara berat badan dan tinggi badan pada balita, baik balita laki-laki maupun perempuan. Artinya, semakin tinggi badan balita cenderung semakin berat pula badannya. Ini merupakan hal yang umum dan sesuai dengan pola pertumbuhan normal pada balita. Grafik *scatter plot* dapat dilihat pada gambar 4.24.

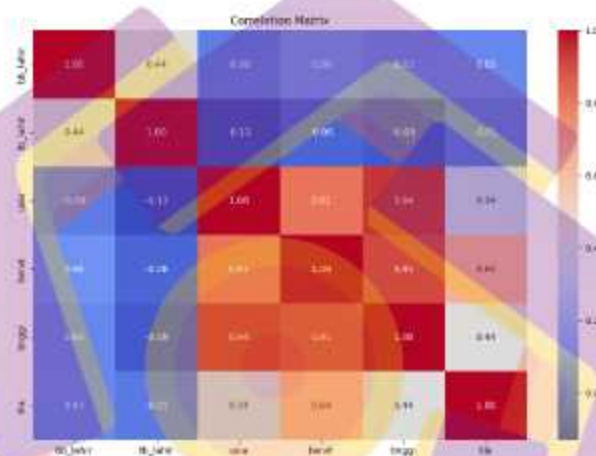


Gambar 4.24. *Scatter plot* berat dan tinggi badan balita

8. Correlation matrix

Matrik korelasi yang menunjukkan hubungan linear antara beberapa beberapa atribut numerik. Setiap kotak pada matriks mewakili korelasi antara dua atribut. Nilai dalam kotak berkisar antara -1 hingga 1, nilai mendekati angka 1 menunjukkan korelasi positif yang kuat, dengan kata lain ketika satu atribut meningkat, atribut lainnya cenderung meningkat pula. Nilai yang menunjukkan angka -1 menunjukkan korelasi negatif yang kuat, artinya ketika satu atribut meningkat, atribut lainnya cenderung menurun, dan nilai yang mendekati 0

menunjukkan tidak ada korelasi yang signifikan. Hasilnya terdapat korelasi positif yang kuat antara atribut usia, tinggi, berat, dan lila. Ini menunjukkan bahwa seiring bertambahnya usia, berat badan, tinggi badan dan lingkaran lengan atas cenderung meningkat, sesuai dengan pertumbuhan fisik balita. Matrik korelasi dapat dilihat 4.25.



Gambar 4.25. Correlation matrix

4.4. Data Preparation

Tahapan *preparation data* dilakukan empat proses, yaitu *label encoding*, *feature importance*, *split data*, dan normalisasi data.

1. Label Encoding

Label encoding adalah teknik dalam *machine learning* dan *preprocessing data* yang mengubah data kategorik yang direpresentasikan dengan label teks menjadi format numerik. Dengan kata lain, teknik ini memberikan nilai numerik yang unik untuk setiap kategori atau label yang terdapat dalam data. Hasil dari

label encoding dengan memberi nilai unik pada setiap kategori dapat dilihat pada gambar 4.26.

	jk	bb_lahir	tb_lahir	usia	berat	tinggi	lila	tb/u	bb/u	bb/tb
0	1	3.4	50.0	30	12.4	86.8	16.0	0	0	0
1	0	2.5	45.0	58	12.8	99.5	15.0	0	1	2
2	1	2.7	47.0	55	14.0	103.8	15.5	0	0	0
3	0	2.8	48.0	54	12.2	96.0	14.0	1	1	2
4	0	2.8	48.0	53	14.2	98.3	16.0	0	0	0

Gambar 4.26. Hasil *label encoding*

Penjelasan terkait hasil dari *label encoding* dengan memberi nilai unik pada setiap kelas dapat dilihat pada tabel 4.2.

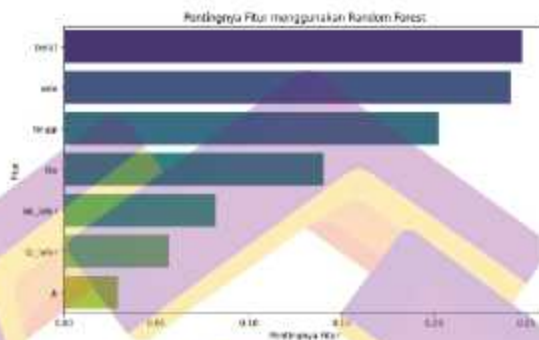
Tabel 4.2. *Label encoding*

Atribut	Kelas	Label Encoding
Jenis kelamin (jk)	L (laki-laki)	0
	P (perempuan)	1
Stunting (tb/u)	Normal	0
	Pendek	1
	Sangat Pendek	2
Wasting (bb/u)	Berat Badan Normal	0
	Kurang	1
	Sangat Kurang	3
Underweight (bb/tb)	Risiko Lebih	2
	Gizi Baik	0
	Gizi Kurang	2
	Risiko Gizi Lebih	5
	Gizi Lebih	3
	Obesitas	4
	Gizi Buruk	1

2. Feature importance

Grafik *feature importance* memberikan gambaran visualisasi mengenai pentingnya setiap fitur dalam memprediksi suatu label, dengan menghitung berdasarkan rata-rata korelasinya. Atribut fitur kemungkinan besar memiliki pengaruh yang signifikan terhadap klasifikasi status gizi balita. Berdasarkan grafik, fitur berat, usia, dan tinggi memiliki pengaruh yang signifikan dalam

memprediksi suatu label. Selanjutnya fitur *lila*, *bb_lahir*, dan *tb_lahir* memiliki pengaruh sedang, sedangkan fitur *jk* memiliki pengaruh yang paling kecil. Grafik *feature importance* dapat dilihat pada gambar 4.27.



Gambar 4.27. *Feature importance*

3. *Split Data*

Dataset dipisah menjadi latih dan data uji untuk memisahkan titik data yang belum diketahui dari model yang sudah dilatih. Tujuannya adalah untuk menguji kinerja model pada data yang tidak pernah dilihat sebelumnya. Proses pemisahan dilakukan secara acak untuk memastikan representasi yang adil dari *dataset*. Data latih digunakan untuk melatih model klasifikasi dalam mengenali karakteristik status gizi balita, seperti *stunting*, *wasting*, dan *underweight*, berdasarkan kelas yang diberikan. Sementara, data uji digunakan untuk mengevaluasi kinerja model yang dilatih. Model akan diterapkan pada data uji yang belum pernah dilihat sebelumnya, dan hasilnya akan dibandingkan dengan nilai sebenarnya untuk mengukur akurasi model.

Dataset dipisahkan menjadi dua variabel terpisah, yaitu variabel X (fitur) dan y (label), dimana X berisi semua fitur dan y mencakup *tb/u*, *bb/u*, dan *bb/tb*.

Data dibagi dengan rasio 80% untuk data latih dan 20 % untuk data uji, menghasilkan data latih sebanyak 2.374 dan 594 data uji. Hasil *split data* dapat dilihat pada gambar 4.28.

Ukuran Data Latih: (2374, 7)
Ukuran Data Uji: (594, 7)

Gambar 4.28. *Split* data latih dan data uji

4. Normalisasi Data

Normalisasi data adalah proses mengubah data numerik ke dalam skala yang sama menggunakan metode standarisasi, proses ini penting untuk memastikan bahwa data yang digunakan untuk melatih dan menguji model memiliki skala yang sama, sehingga model dapat belajar secara efektif. Data yang telah dinormalisasi kemudian akan digunakan dalam tahap pelatihan dan evaluasi model *machine learning*.

4.5. *Modeling Machine Learning*

Klasifikasi multi-label adalah jenis klasifikasi dimana suatu sampel data dapat memiliki lebih dari satu label atau kelas. Berbeda dengan klasifikasi multi-kelas, dimana setiap sampel hanya dapat memiliki satu kelas, dalam multi-label, suatu sampel dapat memiliki kombinasi dari beberapa kelas sekaligus. Klasifikasi yang digunakan pada status gizi balita adalah klasifikasi multi-label. Klasifikasi ini sangat relevan, dimana setiap balita dapat mengalami kondisi gizi pada saat yang sama (simultan), seperti *stunting* (tb/u), *wasting* (bb/u), dan *underweight* (bb/tb).

Dimana kelas pada label tb/u (*stunting*) yaitu normal, pendek, dan sangat pendek. Sedangkan kelas pada label bb/u (*wasting*) yaitu berat badan normal,

kurang, sangat kurang, dan risiko lebih. Dan kelas pada label bb/tb (*underweight*) yaitu gizi baik, gizi kurang, risiko gizi lebih, gizi lebih, obesitas, dan gizi buruk. Berikut adalah contoh data status gizi balita:

Tabel 4.3. Contoh data status gizi

id	jk	bb_lahir	tb_lahir	usia	berat	tinggi	lila	tb/u	bb/u	bb/tb
1	1	3.2	49.0	31	13.2	91.2	15.2	0	0	0
2	0	2.8	49.0	23	8.9	75.0	16.5	2	1	0
3	1	2.8	47.0	51	19.9	105.0	19.0	0	2	5

Pada tabel 4.3, balita pertama berjenis kelamin perempuan dengan status gizi *stunting* kelas normal, *wasting* dengan kelas berat badan normal, dan *underweight* dengan kelas gizi baik. Balita kedua berjenis kelamin laki-laki dengan status gizi *stunting* dengan kelas sangat pendek, *wasting* dengan kelas kurang, dan *underweight* dengan kelas gizi baik. Sedangkan balita ketiga berjenis kelamin perempuan dengan status gizi *stunting* dengan kelas normal, *wasting* dengan kelas risiko lebih, dan *underweight* dengan kelas risiko gizi lebih. Karena setiap balita dapat memiliki kombinasi dari tiga label status gizi tersebut, maka klasifikasi multi-label ini diperlukan untuk mengidentifikasi dan mengklasifikasi kondisi status gizi balita.

Hyperparameter tuning adalah proses dalam *machine learning* yang bertujuan untuk menemukan kombinasi parameter terbaik untuk suatu model. Berbeda dengan model parameter yang dipelajari langsung dari data, *hyperparameter* harus diatur secara manual dan mempengaruhi cara kerja dan kinerja model. Proses ini melibatkan eksplorasi berbagai nilai dan evaluasi kinerja model pada setiap kombinasi. Hal ini disebabkan oleh keragaman algoritma dan parameter yang mempengaruhi kinerja model dalam memprediksi banyak label

secara akurat. Dengan menggunakan teknik *GridSearchCV*, yang secara sistematis mencoba berbagai kombinasi *hyperparameter*, dapat menemukan set parameter paling efektif. *GridSearchCV* melancarkan kombinasi-kombinasi ini berdasarkan seberapa baik kinerja model. Penyesuaian *hyperparameter* yang tepat dapat meningkatkan akurasi klasifikasi label oleh model secara signifikan, sehingga menghasilkan kinerja klasifikasi multi-label yang lebih baik.

Berikut adalah parameter yang digunakan untuk klasifikasi multi-label berdasarkan algoritma yang diterapkan. Parameter dapat dilihat pada tabel 4.4:

Tabel 4.4. Parameter yang digunakan

Algoritma	Parameter
<i>naive bayes (Gaussian)</i>	<ul style="list-style-type: none"> • <code>priors</code>: [[0.5, 0.5], [0.4, 0.6], [0.3, 0.7], [0.2, 0.8], None] • <code>var_smoothing</code>: [1e-1, 1e-2, 1e-3, 1e-4, 1e-5, 1e-6, 1e-7, 1e-8, 1e-9]
<i>logistic regression</i>	<ul style="list-style-type: none"> • <code>C</code>: [0.01, 0.1, 1, 10, 100] • <code>solver</code>: ('lbfgs', 'liblinear', 'newton-cg', 'newton-cholesky', 'sag', 'saga') • <code>penalty</code>: ('l1', 'l2', 'elasticnet', None) • <code>multi_class</code>: ('auto', 'ovr', 'multinomial')
<i>random forest</i>	<ul style="list-style-type: none"> • <code>n_estimators</code>: [50, 100, 200, 300] • <code>max_depth</code>: [[10, 20, 30], None] • <code>min_samples_split</code>: [2, 5, 10] • <code>min_samples_leaf</code>: [1, 2, 4] • <code>max_features</code>: ('sqrt', 'log2', None)
<i>svc</i>	<ul style="list-style-type: none"> • <code>C</code>: [0.1, 1, 10, 100] • <code>kernel</code>: ('linear', 'poly', 'rbf', 'sigmoid') • <code>degree</code>: [3, 4, 5] • <code>gamma</code>: ('scale', 'auto')
<i>decision tree</i>	<ul style="list-style-type: none"> • <code>criterion</code>: ('gini', 'entropy', 'log_loss') • <code>max_depth</code>: [[10, 20, 30], None] • <code>min_samples_split</code>: [2, 5, 10] • <code>min_samples_leaf</code>: [1, 2, 4] • <code>max_features</code>: ('sqrt', 'log2', None)
<i>k-nearest neighbors</i>	<ul style="list-style-type: none"> • <code>n_neighbors</code>: [3, 5, 7, 10, 15] • <code>weights</code>: ('uniform', 'distance') • <code>algorithm</code>: ('auto', 'ball_tree', 'kd_tree', 'brute') • <code>p</code>: [1, 2]


```

Nilai Akurasi Model Logistic Regression untuk Label bb/u: 93.94%
Parameter terbaik untuk Label bb/u: {'C': 0.01, 'multi_class': 'auto', 'penalty': None, 'solver': 'newton-cg'}

Hasil Klasifikasi untuk Label bb/u: [0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 4 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 2 1 0 0 0 0 0 0 0 0 0 0 0 2 0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 1 2 0 0 0 0 0
0 0 0 1 0 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 2 1 0 0 0 2 0 1 0 0 0 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 2 0 0 0 0 1 0 0 0 2 0 1 0 0 0 2 0 0 0 0 0 0 2 1 0 1 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 2 0 0 0 0 0 1 0 0 0 2 0 1 1 0 0 3 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 2 0 2 2 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 3 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 1 0 0 0 0 1 3 3 0 0 0 0 0 0 1 0 0 0 1 0 0 3 0 0 2 0 2 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 2 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 2 2 1 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0]

Laporan Klasifikasi untuk Label bb/u:
              precision    recall  F1-score   support

0               0.87       0.53       0.67         476
1               0.82       0.78       0.80          77
2               0.99       0.98       0.98          34
3               0.54       0.00       0.03          7

accuracy: 0.94
macro avg: 0.81       0.69       0.84        584
weighted avg: 0.94       0.94       0.94        584

```

Gambar 4.33. Hasil klasifikasi model *logistic regression* label bb/u (*wasting*)

Selanjutnya, hasil klasifikasi untuk label bb/tb (*underweight*) menggunakan model *logistic regression* menunjukkan nilai akurasi sebesar 94.78%, dengan parameter terbaik yaitu menggunakan *C: 0.01*, *multi_class: auto*, *penalty: None*, dan *solver: newton-cg*, seperti yang ditunjukkan pada gambar 4.34.

```

Nilai Akurasi Model Logistic Regression untuk Label bb/tb: 94.78%
Parameter terbaik untuk Label bb/tb: {'C': 0.01, 'multi_class': 'auto', 'penalty': None, 'solver': 'newton-cg'}

Hasil Klasifikasi untuk Label bb/tb: [0 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 2 1 0 0 0 2 0 1 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 2 0 0 0 0 1 0 0 0 2 0 1 0 0 0 2 0 0 0 0 0 0 2 1 0 1 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 1 0 0 0 0 1 3 3 0 0 0 0 0 0 1 0 0 0 1 0 0 3 0 0 2 0 2 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 2 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 2 2 1 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0]

Laporan Klasifikasi untuk Label bb/tb:
              precision    recall  F1-score   support

0               0.87       0.99       0.90         589
1               0.88       0.99       0.90           1
2               0.78       0.78       0.78           24
3               0.63       0.77       0.69          13
4               1.00       0.88       0.91           6
5               0.75       0.77       0.76           28

accuracy: 0.95
macro avg: 0.77       0.78       0.71        630
weighted avg: 0.95       0.95       0.95        630

```

Gambar 4.34. Hasil klasifikasi model *log regression* label bb/tb (*underweight*)

min_samples_leaf: 1, *min_samples_split*: 2, dan *n_estimators*: 300, seperti yang ditunjukkan pada gambar 4.36.

```

Nilai Akurasi Model Random Forest untuk label bb/ut 02,89%
Parameter terbaik untuk label bb/ut: {'max_depth': None, 'max_features': None, 'min_samples_leaf': 1, 'min_sampl
Hasil Klasifikasi untuk label bb/ut: [0 0 0 0 0 0 0 1 0 0 0 0 0 0 2 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 2 0 0 1 0 0 0 0 2 0 2 0 0 0
0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 2 0
0 0 1 0 2 2 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
2 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 2
0 2 3 0 0 0 0 1 0 0 2 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 1
0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 2 0 0 0 0 1 0 0 0 2 0 0 1 2 0 2 0 0 0 2 0 2 1 0 1 0 0 0 0 1 0 0 0 1
0 2 0 0 0 0 0 1 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 2 0 0 0 2 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 0 0 0
1 1 0 0 0 1 0 2 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 2 0 0 1 1 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 2 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0
0 0]

Laporan Klasifikasi untuk label bb/ut:

```

	precision	recall	f1-score	support
0	0.99	0.95	0.96	495
1	0.78	0.61	0.74	77
2	0.79	0.79	0.79	58
3	0.67	0.57	0.62	7
accuracy			0.92	594
macro avg	0.89	0.78	0.78	594
weighted avg	0.92	0.62	0.82	594

Gambar 4.36. Hasil klasifikasi model *random forest* label bb/ut (*wasting*)

Selanjutnya, hasil klasifikasi untuk label bb/tb (*underweight*) menggunakan model *random forest* menunjukkan nilai akurasi sebesar 89.73%, dengan parameter terbaik yaitu menggunakan *max_depth*: *None*, *max_features*: *None*, *min_samples_leaf*: 1, *min_samples_split*: 2, dan *n_estimators*: 200, seperti yang ditunjukkan pada gambar 4.37.

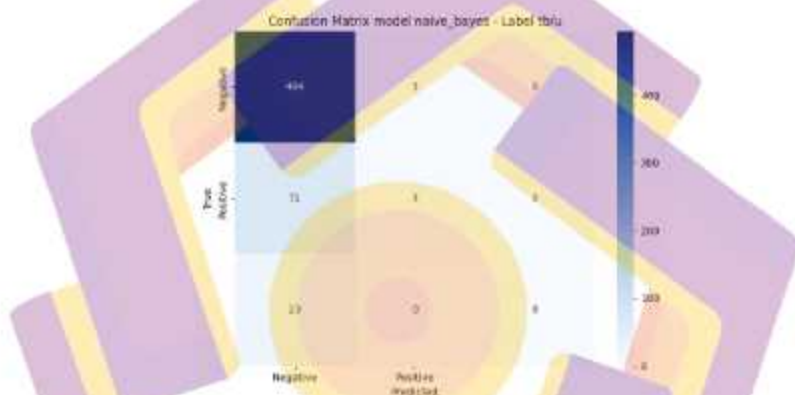
4.6. Evaluasi Model

Setelah pemodelan selesai, langkah selanjutnya adalah menilai kinerja model menggunakan *confusion matrix*. *Confusion matrix* adalah teknik yang digunakan untuk menilai kinerja klasifikasi model. Ini memberikan gambaran visual tentang prediksi yang benar atau salah untuk setiap kelas. *True positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) adalah empat komponen utama matriks ini. TP dan TN menunjukkan jumlah prediksi yang benar, sedangkan FP dan FN menunjukkan kesalahan prediksi. Dengan menggunakan *confusion matrix*, dapat diketahui keakuratan model dan memperoleh metrik penting seperti akurasi, presisi, *recall* dan *f1-score*, yang memberikan gambaran lebih lanjut tentang kekuatan dan kelemahan model.

Berikut penjelasan tabel *confusion matrix* dari model *naïve bayes* label tb/u (*stunting*). Tabel ini membandingkan prediksi yang dibuat oleh model dengan data aktual. Hasilnya menunjukkan bahwa *stunting* pada kelas normal, model berhasil mengklasifikasikan 494 balita sebagai kelas normal dengan benar (TP), namun ada 94 balita kelas normal yang salah diklasifikasikan sebagai kelas lain (FN). Selain itu, terdapat 3 balita yang seharusnya bukan kelas normal namun salah diklasifikasikan sebagai kelas normal (FP).

Selanjutnya *stunting* di kelas pendek, model hanya mampu mengklasifikasikan 3 balita sebagai kelas pendek dengan benar (TP), tetapi ada 94 balita yang seharusnya bukan kelas pendek namun salah diklasifikasikan sebagai kelas pendek (FP). Sebanyak 3 balita kelas pendek juga salah diklasifikasikan sebagai kelas lain (FN).

Selanjutnya *stunting* di kelas sangat pendek, model belum berhasil mengklasifikasikan satupun balita *stunting* sebagai kelas sangat pendek dengan benar (TP). Semua balita dengan kelas sangat pendek justru salah diklasifikasikan sebagai kelas lain (FN), sementara tidak ada kasus yang seharusnya bukan kelas sangat pendek namun salah diklasifikasikan sebagai kelas sangat pendek (FP). *Confusion matrix stunting* model *naïve bayes* dapat dilihat pada gambar 4.47.



Gambar 4.47. *Confusion matrix stunting* (tb/u) model *naïve bayes*

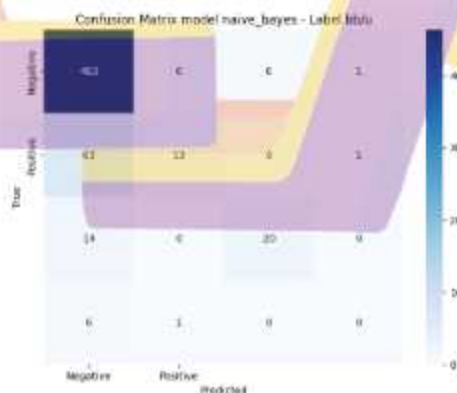
Selanjutnya penjelasan tabel *confusion matrix* dari model *naïve bayes* label bb/u (*wasting*). Hasilnya menunjukkan bahwa *wasting* pada kelas berat badan normal, model berhasil mengklasifikasikan 463 balita sebagai kelas berat badan normal dengan benar (TP), namun ada 83 balita dengan kelas berat badan normal yang salah diklasifikasikan sebagai kelas lain (FN). Selain itu, terdapat 13 balita yang seharusnya bukan kelas berat badan normal namun salah diklasifikasikan sebagai kelas berat badan normal (FP).

Selanjutnya *wasting* di kelas kurang, model berhasil mengklasifikasikan 13 balita sebagai kelas kurang dengan benar (TP), namun ada 7 balita dengan kelas

kurang yang salah diklasifikasikan sebagai kelas lain (FN). Sebanyak 7 balita yang seharusnya bukan kelas kurang juga salah diklasifikasikan sebagai kelas kurang (FP).

Selanjutnya *wasting* di kelas sangat kurang, model berhasil mengklasifikasikan 20 balita sebagai kelas sangat kurang dengan benar (TP), namun ada 6 balita dengan kelas sangat kurang yang salah diklasifikasikan sebagai kelas lain (FN). Sebanyak 6 balita yang seharusnya bukan kelas sangat kurang juga salah diklasifikasikan sebagai kelas sangat kurang (FP).

Selanjutnya *wasting* kelas risiko lebih, model belum berhasil mengklasifikasikan satupun balita *wasting* sebagai kelas risiko lebih dengan benar (TP). Semua balita dengan risiko lebih justru salah diklasifikasikan sebagai kelas lain (FN), sementara hanya ada 2 balita yang seharusnya bukan kelas risiko lebih namun salah diklasifikasikan sebagai kelas risiko lebih (FP). *Confusion matrix* *wasting* model *naïve bayes* dapat dilihat pada gambar 4.48.



Gambar 4.48. *Confusion matrix* *wasting* (bb/u) model *naïve bayes*

Selanjutnya penjelasan tabel *confusion matrix* dari model *naïve bayes* label bb/tb (*underweight*). Hasilnya menunjukkan bahwa *underweight* pada kelas gizi baik, model ini cukup baik dalam mengklasifikasikan 499 balita sebagai kelas gizi baik (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain (FN) maupun yang seharusnya bukan kelas gizi baik namun diklasifikasikan sebagai kelas gizi baik (False Positive).

Selanjutnya *underweight* di kelas gizi buruk, gizi kurang, gizi lebih, obesitas, dan risiko gizi lebih, model menunjukkan kesulitan dalam mengklasifikasikan dengan akurat individu pada kelas ini. Terdapat banyak balita *underweight* yang salah diklasifikasikan, baik sebagai kelas yang benar maupun kelas lain. Hal ini mengindikasikan bahwa model masih perlu ditingkatkan performanya, terutama dalam membedakan antara kelas-kelas yang memiliki karakteristik yang mirip. *Confusion matrix underweight* model *naïve bayes* dapat dilihat pada gambar 4.49.

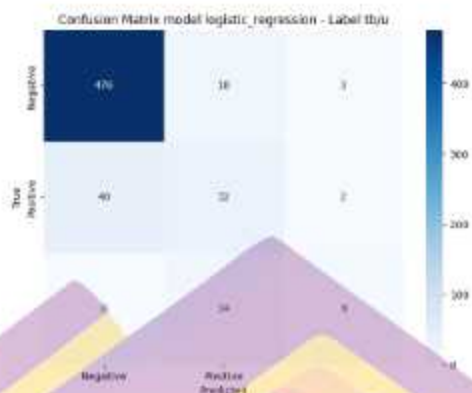


Gambar 4.49. *Confusion matrix underweight* (bb/tb) model *naïve bayes*

Berikut penjelasan tabel *confusion matrix* dari model *logistic regression* label tb/u (*stunting*). Tabel ini membandingkan prediksi yang dibuat oleh model dengan data aktual. Hasilnya menunjukkan bahwa *stunting* pada kelas normal, model berhasil mengklasifikasikan 476 balita sebagai kelas normal dengan benar (TP), namun ada 21 balita dengan kelas normal yang salah diklasifikasikan sebagai kelas lain (FN). Selain itu, terdapat 40 balita yang seharusnya bukan kelas normal namun salah diklasifikasikan sebagai kelas normal (FP).

Selanjutnya *stunting* di kelas pendek, model berhasil mengklasifikasikan 32 balita sebagai kelas pendek dengan benar (TP), namun ada 42 balita dengan kelas pendek yang salah diklasifikasikan sebagai kelas lain (FN). Sebanyak 32 balita yang seharusnya bukan kelas pendek juga salah diklasifikasikan sebagai kelas pendek (FP).

Selanjutnya *stunting* di kelas sangat pendek, model berhasil mengklasifikasikan 9 balita sebagai kelas sangat pendek dengan benar (TP), namun ada 14 balita dengan kelas sangat pendek yang salah diklasifikasikan sebagai kelas lain (FN). Sebanyak 5 balita yang seharusnya bukan kelas sangat pendek juga salah diklasifikasikan sebagai kelas sangat pendek (FP). *Confusion matrix stunting* model *logistic regression* dapat dilihat pada gambar 4.50.



Gambar 4.50. *Confusion matrix stunting (tb/u) model logistic regression*

Selanjutnya penjelasan tabel *confusion matrix* dari model *logistic regression* label bb/u (*wasting*). Hasilnya menunjukkan bahwa *wasting* pada kelas berat badan normal, model berhasil mengklasifikasikan 460 balita sebagai kelas berat badan normal dengan benar (TP), namun ada 16 balita dengan kelas berat badan normal yang salah diklasifikasikan sebagai kelas lain (FN). Selain itu, terdapat 15 balita yang seharusnya bukan kelas berat badan normal namun salah diklasifikasikan sebagai kelas berat badan normal (FP).

Selanjutnya *wasting* di kelas kurang, model berhasil mengklasifikasikan 60 balita sebagai kelas kurang dengan benar (TP), namun ada 17 balita dengan kelas kurang yang salah diklasifikasikan sebagai kelas lain (FN). Sebanyak 13 balita yang seharusnya bukan kelas kurang juga salah diklasifikasikan sebagai kelas kurang (FP).

Selanjutnya *wasting* di kelas sangat kurang, model berhasil mengklasifikasikan 32 balita sebagai kelas sangat kurang dengan benar (TP), namun ada 2 balita dengan kelas sangat kurang yang salah diklasifikasikan sebagai

kelas lain (FN). Sebanyak 2 balita yang seharusnya bukan kelas sangat kurang juga salah diklasifikasikan sebagai kelas sangat kurang (FP).

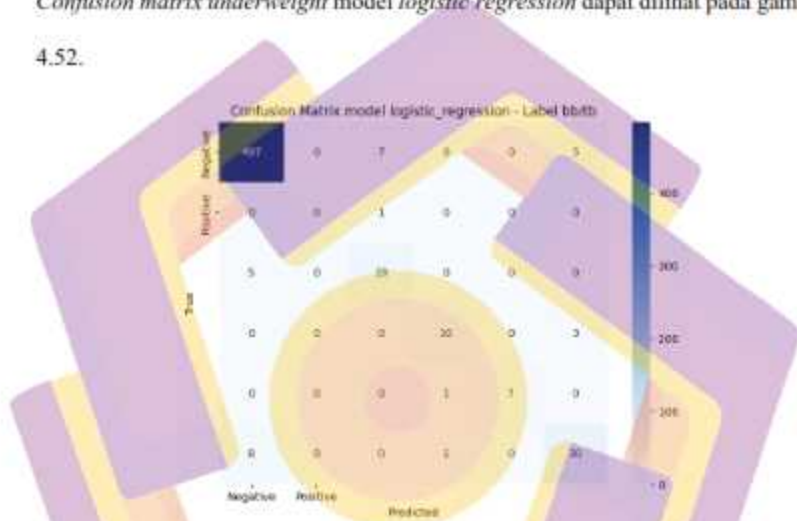
Selanjutnya *wasting* di kelas risiko lebih, model berhasil mengklasifikasikan 6 balita sebagai kelas risiko lebih dengan benar (TP), namun ada 1 balita dengan kelas risiko lebih yang salah diklasifikasikan sebagai kelas lain (FN). Sebanyak 6 balita yang seharusnya bukan kelas risiko lebih juga salah diklasifikasikan sebagai kelas risiko lebih (FP). *Confusion matrix* *wasting* model *logistic regression* dapat dilihat pada gambar 4.51.



Gambar 4.51. *Confusion matrix* *wasting* (bb/u) model *logistic regression*

Selanjutnya penjelasan tabel *confusion matrix* dari model *logistic regression* label bb/tb (*underweight*). Hasilnya menunjukkan bahwa *underweight* pada kelas gizi baik, model berhasil mengklasifikasikan sebagian besar sebanyak 497 balita sebagai kelas gizi baik (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain (FN) maupun yang seharusnya bukan kelas gizi baik namun diklasifikasikan sebagai gizi baik (FP).

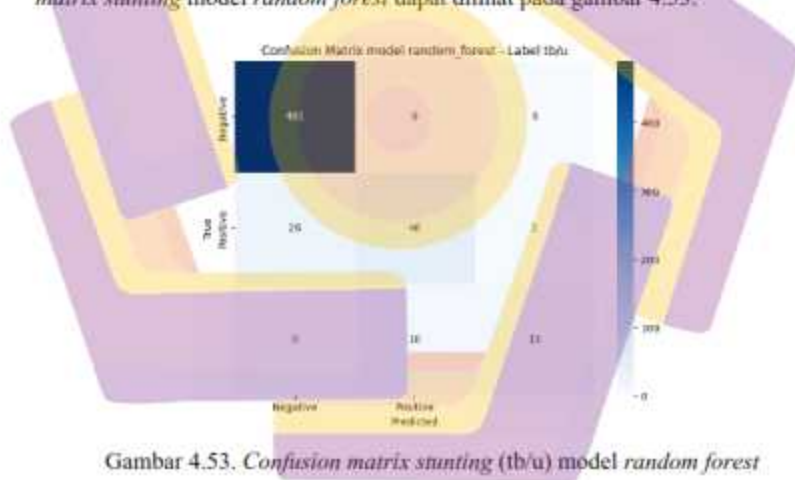
Selanjutnya *underweight* di kelas gizi buruk, gizi kurang, gizi lebih, obesitas, dan risiko gizi lebih, model menunjukkan kinerja yang bervariasi pada kelas-kelas ini. Beberapa kelas memiliki tingkat akurasi yang cukup baik, sementara kelas lainnya masih memiliki tingkat kesalahan yang cukup tinggi. *Confusion matrix underweight* model *logistic regression* dapat dilihat pada gambar 4.52.



Gambar 4.52. *Confusion matrix underweight* (bb/tb) model *logistic regression*

Berikut penjelasan tabel *confusion matrix* dari model *random forest* label *tb/u* (*stunting*). Tabel ini membandingkan prediksi yang dibuat oleh model dengan data aktual. Hasilnya menunjukkan bahwa *stunting* pada kelas normal, model berhasil mengklasifikasikan 491 balita sebagai kelas normal dengan benar (TP), namun ada 6 balita dengan kelas kelas normal yang salah diklasifikasikan sebagai kelas lain (FN). Selain itu, terdapat 26 balita yang seharusnya bukan kelas normal namun salah diklasifikasikan sebagai kelas normal (FP).

Selanjutnya *stunting* di kelas pendek, model berhasil mengklasifikasikan 46 balita sebagai kelas pendek dengan benar (TP), namun ada 28 balita dengan kelas pendek yang salah diklasifikasikan sebagai kelas lain (FN). Sebanyak 16 balita yang seharusnya bukan kelas pendek juga salah diklasifikasikan sebagai kelas pendek (FP). Selanjutnya *stunting* di kelas sangat pendek, model berhasil mengklasifikasikan 13 balita sebagai kelas sangat pendek dengan benar (TP), namun ada 10 balita dengan kelas sangat pendek yang salah diklasifikasikan sebagai kelas lain (FN). Sebanyak 2 balita yang seharusnya bukan kelas sangat pendek juga salah diklasifikasikan sebagai kelas sangat pendek (FP). *Confusion matrix stunting* model *random forest* dapat dilihat pada gambar 4.53.

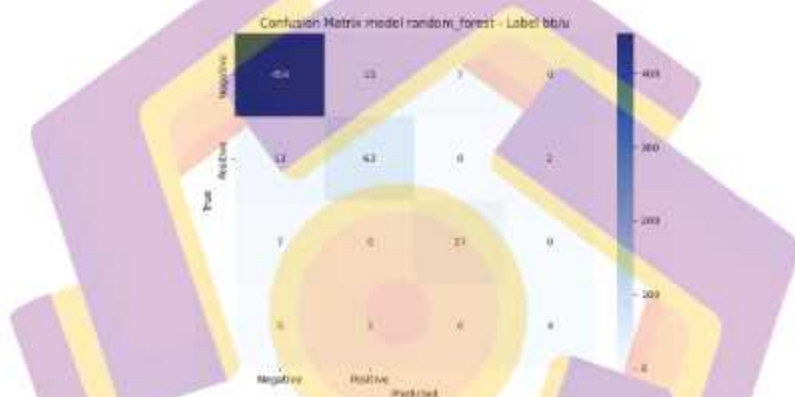


Gambar 4.53. *Confusion matrix stunting (tb/u) model random forest*

Selanjutnya penjelasan tabel *confusion matrix* dari model *random forest* label bb/u (*wasting*). Hasilnya menunjukkan bahwa *wasting* pada kelas berat badan normal, model berhasil mengklasifikasikan sebagian besar sebanyak 454 balita sebagai kelas berat badan normal (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain (FN) maupun yang

seharusnya bukan kelas berat badan normal namun diklasifikasikan sebagai kelas berat badan normal (FP).

Selanjutnya *wasting* di kelas kurang, sangat kurang, dan risiko lebih, model juga menunjukkan kinerja yang cukup baik dalam mengklasifikasikan balita pada kelas ini, meskipun masih terdapat beberapa kesalahan klasifikasi. *Confusion matrix wasting* model *random forest* dapat dilihat pada gambar 4.54.

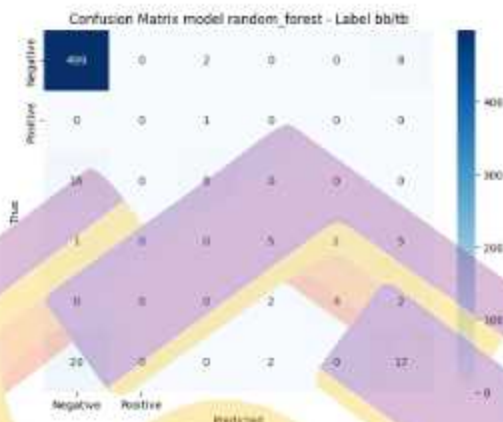


Gambar 4.54. *Confusion matrix wasting (bb/u) model random forest*

Selanjutnya penjelasan tabel *confusion matrix* dari model *random forest* label bb/tb (*underweight*). Hasilnya menunjukkan bahwa *underweight* pada kelas gizi baik, model berhasil mengklasifikasikan sebagian besar sebanyak 499 balita sebagai kelas gizi baik (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain (FN) maupun yang seharusnya bukan kelas gizi baik namun diklasifikasikan kelas gizi baik (FP).

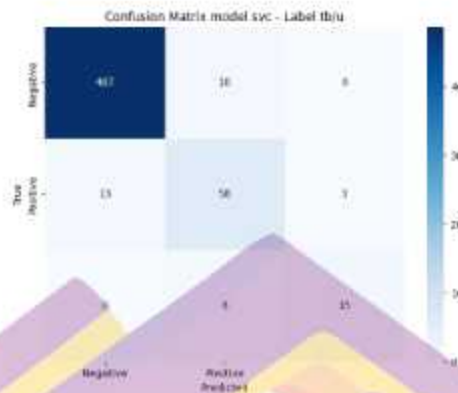
Selanjutnya *underweight* di kelas gizi buruk, gizi kurang, gizi lebih, obesitas, dan risiko gizi lebih, model juga menunjukkan kinerja yang cukup baik dalam mengklasifikasikan status gizi balita pada kelas ini, meskipun masih terdapat

beberapa kesalahan klasifikasi. *Confusion matrix underweight* model *random forest* dapat dilihat pada gambar 4.55.



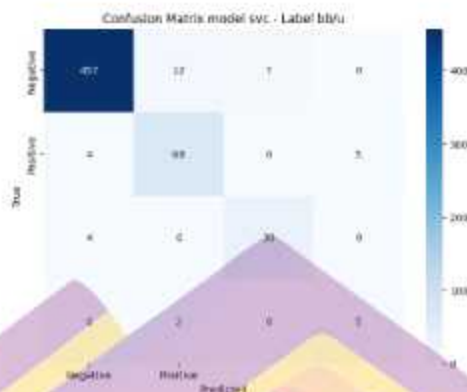
Gambar 4.55. *Confusion matrix underweight* (bb/tb) model *random forest*

Berikut penjelasan tabel *confusion matrix* dari model *svc* label *tb/u* (*stunting*). Tabel ini membandingkan prediksi yang dibuat oleh model dengan data aktual. Hasilnya menunjukkan bahwa *stunting* pada kelas normal, model berhasil mengklasifikasikan sebagian besar, sebanyak 487 balita sebagai kelas normal (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain (FN) maupun yang seharusnya bukan kelas normal namun diklasifikasikan sebagai kelas normal (FP). Selanjutnya *stunting* di kelas pendek dan sangat pendek, model juga menunjukkan kinerja yang cukup baik dalam mengklasifikasikan status gizi balita pada kelas ini, meskipun masih terdapat beberapa kesalahan klasifikasi. *Confusion matrix stunting* model *svc* dapat dilihat pada gambar 4.56.



Gambar 4.56. *Confusion matrix stunting (tb/u) model svc*

Selanjutnya penjelasan tabel *confusion matrix* dari model *svc* label *bb/u* (*wasting*). Hasilnya menunjukkan bahwa *wasting* pada kelas berat badan normal, model berhasil mengklasifikasikan sebagian besar, sebanyak 457 balita sebagai kelas berat badan normal (TP), dengan sedikit kesalahan klasifikasi baik sebagai kelas lain (FN) maupun sebaliknya (FP). Selanjutnya *wasting* di kelas kurang, sangat kurang, dan risiko lebih, model juga menunjukkan kinerja yang cukup baik dalam mengklasifikasikan status gizi balita pada kelas ini, meskipun masih terdapat beberapa kesalahan klasifikasi. *Confusion matrix wasting* model *svc* dapat dilihat pada gambar 4.57.



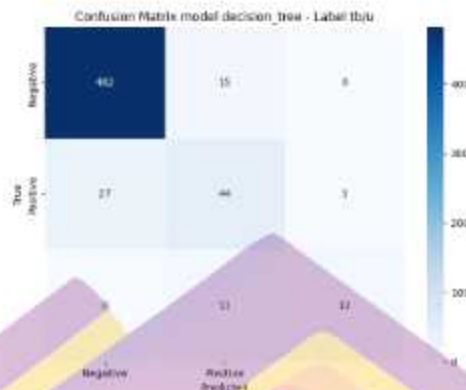
Gambar 4.57. *Confusion matrix wasting (bb/u) model svc*

Selanjutnya penjelasan tabel *confusion matrix* dari model *svc* label bb/tb (*underweight*). Hasilnya menunjukkan bahwa *underweight* kelas gizi baik, model berhasil mengklasifikasikan sebagian besar sebanyak 498 balita sebagai kelas gizi baik (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain (FN) maupun yang seharusnya bukan kelas gizi baik namun diklasifikasikan kelas gizi baik (FP). Selanjutnya *underweight* di kelas gizi buruk, gizi kurang, gizi lebih, obesitas, dan risiko gizi lebih, model juga menunjukkan kinerja yang cukup baik dalam mengklasifikasikan status gizi balita pada kelas ini, meskipun masih terdapat beberapa kesalahan klasifikasi. *Confusion matrix underweight* model *svc* dapat dilihat pada gambar 4.58.



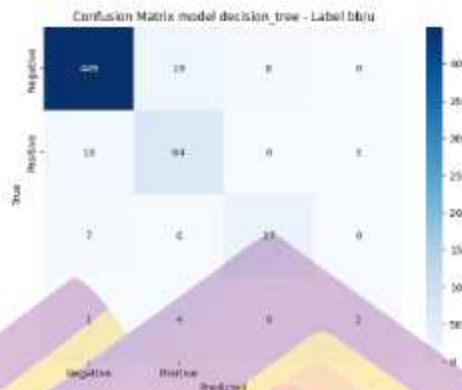
Gambar 4.58. *Confusion matrix underweight (bb/tb) model svc*

Berikut penjelasan tabel *confusion matrix* dari model *decision tree* label *tb/u* (*stunting*). Tabel ini membandingkan prediksi yang dibuat oleh model dengan data aktual. Hasilnya menunjukkan bahwa *stunting* pada kelas normal, model berhasil mengklasifikasikan sebagian besar, sebanyak 482 balita kelas normal (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain (FN) maupun yang seharusnya bukan kelas normal namun diklasifikasikan sebagai kelas normal (FP). Selanjutnya *stunting* di kelas pendek dan sangat pendek, model juga menunjukkan kinerja yang cukup baik dalam mengklasifikasikan status gizi balita pada kelas ini, meskipun masih terdapat beberapa kesalahan klasifikasi. *Confusion matrix stunting model decision tree* dapat dilihat pada gambar 4.59.



Gambar 4.59. *Confusion matrix stunting (tb/u) model decision tree*

Selanjutnya penjelasan tabel *confusion matrix* dari model *decision tree* label bb/u (*wasting*). Hasilnya menunjukkan bahwa *wasting* pada kelas berat badan normal, model berhasil mengklasifikasikan sebagian besar, sebanyak 449 balita sebagai kelas berat badan normal (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain (FN) maupun yang seharusnya bukan kelas berat badan normal namun diklasifikasikan sebagai kelas berat badan normal (FP). Selanjutnya *wasting* di kelas kurang, sangat kurang, dan risiko lebih, model juga menunjukkan kinerja yang cukup baik dalam mengklasifikasikan status gizi balita pada kelas ini, meskipun masih terdapat beberapa kesalahan klasifikasi, terutama pada kelas risiko lebih yang memiliki jumlah data yang lebih sedikit. *Confusion matrix wasting* model *decision tree* dapat dilihat pada gambar 4.60.



Gambar 4.60. *Confusion matrix wasting (bb/u) model decision tree*

Selanjutnya penjelasan tabel *confusion matrix* dari model *decision tree* label bb/tb (*underweight*). Hasilnya menunjukkan bahwa *underweight* pada kelas gizi baik, model berhasil mengklasifikasikan sebagian besar, sebanyak 482 balita kelas gizi baik (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain (FN) maupun yang seharusnya bukan kelas gizi baik namun diklasifikasikan sebagai kelas gizi baik (FP). Selanjutnya *underweight* di kelas gizi buruk, gizi kurang, gizi lebih, obesitas, dan risiko gizi lebih, model juga menunjukkan kinerja yang cukup baik dalam mengklasifikasikan status gizi balita pada kelas ini, meskipun masih terdapat beberapa kesalahan klasifikasi. *Confusion matrix underweight model decision tree* dapat dilihat pada gambar 4.61.



Gambar 4.62. *Confusion matrix stunting (tb/u) model k-nearest neighbors*

Selanjutnya penjelasan tabel *confusion matrix* dari model *k-nearest neighbors* label bb/u (*wasting*). Hasilnya menunjukkan bahwa *wasting* pada kelas berat badan normal, model berhasil mengklasifikasikan sebagian besar, sebanyak 458 balita sebagai kelas berat badan normal (TP), namun masih terdapat beberapa kesalahan klasifikasi, baik yang diklasifikasikan sebagai kelas lain sebanyak 69 balita (FN) maupun yang seharusnya bukan kelas berat badan normal namun diklasifikasikan sebagai kelas berat badan normal sebanyak 18 balita (FP).

Selanjutnya *wasting* di kelas kurang, sangat kurang, dan risiko lebih, model mengalami kesulitan dalam mengklasifikasikan status gizi balita pada kelas ini dengan tingkat akurasi yang lebih rendah dibandingkan kelas berat badan normal, ditunjukkan oleh nilai TP yang lebih rendah dan FP serta FN yang lebih tinggi. *Confusion matrix wasting* model *k-nearest neighbors* dapat dilihat pada gambar 4.63.



Gambar 4.63. *Confusion matrix wasting (bb/u) model k-nearest neighbors*

Selanjutnya penjelasan tabel *confusion matrix* dari model *k-nearest neighbors* label bb/tb (*underweight*). Hasilnya menunjukkan bahwa *underweight* pada kelas gizi baik, model berhasil mengklasifikasikan sebagian besar, sebanyak 505 balita sebagai kelas gizi baik (TP), namun masih terdapat kesalahan klasifikasi yang cukup signifikan, terutama pada FN yang tinggi sebanyak 59 balita. Ini mengindikasikan banyak balita yang seharusnya diklasifikasikan sebagai kelas lain seperti, gizi kurang namun salah diklasifikasikan sebagai gizi baik.

Selanjutnya *underweight* di kelas gizi buruk, gizi kurang, gizi lebih, obesitas, dan risiko gizi lebih, model mengalami kesulitan dalam mengklasifikasikan status gizi balita pada kelas ini dengan tingkat akurasi yang rendah, ditunjukkan oleh nilai TP yang rendah dan FP serta FN yang tinggi. *Confusion matrix underweight* model *k-nearest neighbors* dapat dilihat pada gambar 4.64.



Gambar 4.64. Confusion matrix underweight (bb/tb) model *k-nearest neighbors*

4.7. Testing

Tujuan dari pengujian ini adalah untuk mengevaluasi kinerja model *machine learning* yang digunakan dalam klasifikasi terhadap hasil klasifikasi dengan data yang sebenarnya. Dapat dilihat seberapa akurat model tersebut dalam mengklasifikasi data ke dalam kelas benar. Pada *testing* model *naïve bayes* terdapat perbedaan nilai antara nilai sebenarnya dengan hasil klasifikasi pada gambar 4.65.

Hasil untuk model *naïve bayes*:

	y_test_bb/u	y_pred_naive_bayes_bb/u	y_test_bb/tb	y
2778	1	0	0	1
1415	0	0	0	0
2775	0	0	0	0
388	1	0	0	1
1696	0	0	0	0

	y_pred_naive_bayes_bb/u	y_test_bb/tb	y_pred_naive_bayes_bb/tb
2778	0	0	0
1415	0	0	0
2775	0	0	0
388	0	0	0
1696	0	0	0

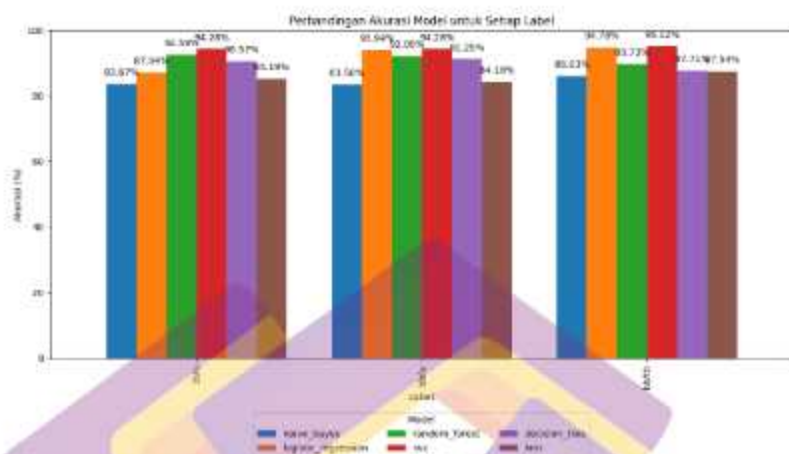
Gambar 4.65. Hasil pengujian model *naïve bayes*

	Model_Label	Accuracy
11	svc_bb/tb	95.12%
5	logistic_regression_bb/tb	94.78%
9	svc_tb/u	94.28%
10	svc_bb/u	94.28%
4	logistic_regression_bb/u	93.94%
6	random_forest_tb/u	92.59%
7	random_forest_bb/u	92.09%
13	decision_tree_bb/u	91.25%
12	decision_tree_tb/u	90.57%
8	random_forest_bb/tb	89.73%
14	decision_tree_bb/tb	87.71%
17	knn_bb/tb	87.54%
3	logistic_regression_tb/u	87.04%
2	naive_bayes_bb/tb	86.03%
15	knn_tb/u	85.19%
16	knn_bb/u	84.18%
0	naive_bayes_tb/u	83.67%
1	naive_bayes_bb/u	83.58%

Gambar 4.77. Perbandingan model akurasi terbaik

Hasil eksperimen pengujian menunjukkan variasi akurasi model dalam melakukan klasifikasi label kelas pada data uji. Dari hasil perbandingan, terlihat bahwa model *svc* pada urutan pertama dan *logistic regression* pada urutan kedua secara konsisten memberikan akurasi yang lebih tinggi. Model *svc* mencatat nilai akurasi tb/u 94.28%, 94.28% untuk bb/u, dan 95.12% untuk bb/tb, diikuti model *logistic regression* dengan nilai akurasi sebesar 87.04% untuk tb/u, 93.94% untuk bb/u, dan 94.78% untuk bb/tb.

Selanjutnya model *random forest* memperoleh nilai akurasi sebesar 92.59% untuk tb/u, 92.09% untuk bb/u, dan 89.73% untuk bb/tb. Model *decision tree* menunjukkan nilai akurasi sebesar 90.57% untuk tb/u, 91.25% untuk bb/u, dan 87.71% untuk bb/tb. Kemudian model *k-nearest neighbors* menunjukkan nilai akurasi sebesar 85.19% untuk tb/u, 84.18% untuk bb/u, dan 87.54% untuk bb/tb. Di urutan terakhir, model *naive bayes* memperoleh nilai akurasinya yaitu, sebesar 83.67% untuk tb/u, 83.50% untuk bb/u, dan 86.03% untuk bb/tb. Grafik perbandingan model dengan nilai akurasinya dapat dilihat pada gambar 4.78.



Gambar 4.78. Perbandingan akurasi model *machine learning* setiap label

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa:

1. Berdasarkan hasil penelitian, algoritma *machine learning* yang memiliki tingkat akurasi tertinggi dalam mengklasifikasi multi-label status gizi balita adalah *Support Vector Classifier (SVC)*. Algoritma ini secara konsisten menunjukkan performa terbaik dalam mengklasifikasikan berbagai label status gizi balita, dengan akurasi 94.28% untuk *stunting* (tb/u), 94.28% untuk *wasting* (bb/u), dan 95.12% untuk *underweight* (bb/tb), lebih tinggi dibanding dengan algoritma lainnya.
2. Kombinasi *hyperparameter* yang optimal untuk setiap algoritma berkontribusi secara signifikan terhadap peningkatan akurasi dalam klasifikasi multi-label status gizi balita. Melalui proses *tuning hyperparameter*, ditemukan bahwa pengaturan optimal pada algoritma *svc* menghasilkan akurasi tertinggi. Untuk label tb/u (*stunting*), kombinasi parameter terbaik adalah $C: 100$, $degree: 3$, $gamma: scale$, dan $kernel: rbf$. Untuk label bb/u (*wasting*), kombinasi parameter terbaik adalah $C: 100$, $degree: 3$, $gamma: auto$, dan $kernel: rbf$. Sedangkan untuk label bb/tb (*underweight*), kombinasi parameter terbaik adalah $C: 100$, $degree: 3$, $gamma: scale$, dan $kernel: linear$. Hasil ini menegaskan

pentingnya proses *hyperparameter tuning* dalam memperoleh hasil klasifikasi yang lebih akurat dan andal.

5.2. Saran

Berdasarkan kesimpulan yang diperoleh, disarankan agar algoritma *svc* diterapkan lebih luas dalam aplikasi yang memerlukan deteksi atau pemantauan status gizi balita, mengingat akurasinya yang tertinggi dalam klasifikasi multi-label. Selain itu, disarankan untuk melakukan eksplorasi lebih mendalam terhadap metode optimasi *hyperparameter* lainnya, seperti *Bayesian Optimization* atau *Genetic Algorithms*, guna mencari kemungkinan peningkatan akurasi lebih lanjut. Pengujian model *svc* dengan kombinasi *hyperparameter* optimal pada *dataset* yang lebih luas dan beragam juga penting untuk memastikan generalisasi dan keandalan model. Selain itu, disarankan pula mengembangkan sistem prediksi status gizi berbasis *svc* yang dapat digunakan oleh tenaga kesehatan di lapangan, disertai pelatihan agar dapat memahami dan dapat menginterpretasikan hasil klasifikasi dengan efektif di lapangan.

DAFTAR PUSTAKA

- Adhinata, F. D., & Junaidi, A. (2022). Gender Classification on Video Using FaceNet Algorithm and Supervised Machine Learning. *Deanship of Graduate Studies and Scientific Research, University of Bahrain*, 11(1).
- Aditya, M. F. R., Azizah, N. L., & Indahyanti, U. (2024). Prediksi Penyakit Hipertensi Menggunakan Metode Decision Tree dan Random Forest. *JURNAL ILMIAH KOMPUTER & SISTEM INFORMASI*, 23(1), 9–15. <https://doi.org/10.32409/jikstik.23.1.3503>
- Afiatuddin, N., Wicaksono, M. T., Akbar, V. R., Rahmaddeni, & Wulandari, D. (2024). Komparasi Algoritma Machine Learning dalam Klasifikasi Kanker Payudara. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(2), 889–899. <https://doi.org/10.30865/mib.v8i2.7457>
- Ali, A., & Mashwani, W. K. (2023). A Supervised Machine Learning Algorithms: Applications, Challenges, and Recommendations. *Proceedings of the Pakistan Academy of Sciences: Part A*, 60(4), 1–12. [https://doi.org/10.53560/PPASA\(60-4\)831](https://doi.org/10.53560/PPASA(60-4)831)
- Annasaheb, A. B., & Verma, V. K. (2016). Data Mining Classification Techniques: A Recent Survey. *International Journal of Emerging Technologies in Engineering Research (IJETER)*, 4(8). www.ijeter.everscience.org
- Aritonang, I. (2013). *Memantau dan Menilai Status Gizi Anak*. Leutika Books.
- Chilyabanyama, O. N., Chilengi, R., Simuyandi, M., Chisenga, C. C., Chirwa, M., Hamusonde, K., Saroj, R. K., Iqbal, N. T., Ngaruye, I., & Bosomprah, S. (2022). Performance of Machine Learning Classifiers in Classifying Stunting

- among Under-Five Children in Zambia. *Children*, 9(7), 1–11.
<https://doi.org/10.3390/children9071082>
- Darnila, E., Maryana, & Azmi, M. (2021). APLIKASI KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN METODE NAÏVE BAYES BERBASIS ANDROID. *METHOMIKA: Jurnal Manajemen Informatika dan Komputersasi Akuntansi*, 5(2), 135–141.
<https://doi.org/10.46880/jmika.Vol5No2.pp135-141>
- Diantika, A. S., & Firmanto, Y. (2020). IMPLEMENTASI MACHINE LEARNING PADA APLIKASI PENJUALAN PRODUK DIGITAL (STUDI PADA GRABKIOS). <https://jimfeb.ub.ac.id/index.php/jimfeb/article/view/6928>
- Firdausi, N., Wessiani, N. A., & Santosa, B. (2012). ANALISIS FINANCIAL DISTRESS DENGAN PENDEKATAN DATA MINING PADA INDUSTRI MANUFAKTUR GO-PUBLIC DI INDONESIA. http://www.ie.its.ac.id/rbti/index.php?p=show_detail&id=6028
- Ho, J. Y., Afan, H. A., El-Shafie, A. H., Koting, S. B., Mohd, N. S., Jaafar, W. Z. B., Lai Sai, H., Malek, M. A., Ahmed, A. N., Mohtar, W. H. M. W., Elshorbagy, A., & El-Shafie, A. (2019). Towards a time and cost effective approach to water quality index class prediction. *Journal of Hydrology*, 575, 148–165. <https://doi.org/10.1016/j.jhydrol.2019.05.016>
- Isnaini, N., Adiwijaya, Mubarak, M. S., & Bakar, M. Y. A. (2019). A multi-label classification on topics of Indonesian news using K-Nearest Neighbor. *Journal of Physics: Conference Series*, 1192, 012027. <https://doi.org/10.1088/1742-6596/1192/1/012027>

- Istiany, A., & Rusilanti. (2013). *Gizi Terapan*. Remaja Rosdakarya.
- Kementerian Kesehatan RI. (2023, Januari 25). *Prevalensi Stunting di Indonesia Turun ke 21,6% dari 24,4%*.
<https://www.kemkes.go.id/article/view/23012500002/prevalensi-stunting-di-indonesia-turun-ke-21-6-dari-24-4-.html>
- Kementerian Kesehatan RI, B. K. P. K. (2022). *BUKU SAKU Hasil Survei Status Gizi Indonesia (SSGI) 2022*.
<https://kesmas.kemkes.go.id/assets/uploads/contents/attachments/09fb5b8ccfd088080f2521ff0b4374f.pdf>
- Khonsa Izzaty, A. M., Mubarak, M. S., Huda, N. S., & Adiwijaya. (2018). A Multi-Label Classification on Topics of Quranic Verses in English Translation Using Tree Augmented Naïve Bayes. *2018 6th International Conference on Information and Communication Technology (ICoICT)*, 103–106.
<https://doi.org/10.1109/ICoICT.2018.8528802>
- Kumar, V., Lalotra, G. S., & Kumar, R. K. (2022). Improving performance of classifiers for diagnosis of critical diseases to prevent COVID risk. *Computers and Electrical Engineering*, 102.
<https://doi.org/10.1016/j.compeleceng.2022.108236>
- Laela, I. N., & Baihaqi, W. M. (2024). Analisis Klasifikasi Penyakit Multiple Sclerosis Menggunakan Algoritma Logistic Regression dan SVM. *GENERATION JOURNAL*, 8(1), 2580–4952.
<https://doi.org/10.29407/gj.v8i1.20646>

- Lonang, S., Yudhana, A., & Biddinika, M. K. (2023). Performance Analysis for Classification of Malnourished Toddlers Using K-Nearest Neighbor. *Scientific Journal of Informatics*, 10(3), 313–322. <https://doi.org/10.15294/sji.v10i3.45196>
- Lonang, S., Yudhana, A., & Kunta Biddinika, M. (2023). Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting. *Jurnal Media Informatika Budidarma*, 7(4), 2109–2117. <https://doi.org/10.30865/mib.v7i4.6553>
- Neelamegam, S., & Ramaraj, Dr. E. (2013). Classification algorithm in Data mining: An Overview. *International Journal of P2P Network Trends and Technology*, 3(5). <https://ijpttjournal.org/archives/ijptt-v3i8p101>
- Nofriansyah, D., & Nurcahyo, G. W. (2015). *Algoritma Data Mining Dan Pengujian*. Deepublish. https://books.google.co.id/books?id=Fn-QDwAAQBAJ&hl=id&source=gbs_navlinks_s
- Noor, N. Bin, Anwar, Md. S., & Dey, M. (2019). Comparative Study Between Decision Tree, SVM and KNN to Predict Anaemic Condition. *IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health*.
- Pane, R. A., Mubarak, M. S., Huda, N. S., & Adiwijaya. (2018). A Multi-Lable Classification on Topics of Quranic Verses in English Translation Using Multinomial Naive Bayes. *2018 6th International Conference on Information and Communication Technology (ICoICT)*, 481–484. <https://doi.org/10.1109/ICoICT.2018.8528777>

- Pertiwi, A. G., Bachtiar, N., Kusumaningrum, R., Waspada, I., & Wibowo, A. (2020). Comparison of performance of k-nearest neighbor algorithm using smote and k-nearest neighbor algorithm without smote in diagnosis of diabetes disease in balanced data. *Journal of Physics: Conference Series*, 1524(1), 012048. <https://doi.org/10.1088/1742-6596/1524/1/012048>
- Rahman, S. M. J., Ahmed, N. A. M. F., Abedin, M. M., Ahammed, B., Ali, M., Rahman, M. J., & Md Maniruzzaman. (2021). Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. *PLoS ONE*, 16(6 June 2021), 1–11. <https://doi.org/10.1371/journal.pone.0253172>
- Rizqullah, M. F., Raihana, N. T., & Jambak, M. I. (2024). Komparasi Penerapan Algoritma C4.5, K-Nearest Neighbor, dan Naïve Bayes untuk Keberlangsungan Pasien Gagal Jantung. *KLIK: KAJIAN ILMIAH INFORMATIKA DAN KOMPUTER*, 4(5), 2580–2587. <https://doi.org/10.30865/klik.v4i5.1788>
- Sahu, B., Mohanty, S. N., & Rout, S. K. (2019). A Hybrid Approach for Breast Cancer Classification and Diagnosis. *EAI Endorsed Transactions on Scalable Information Systems*, 6(20). <https://doi.org/10.4108/eai.19-12-2018.156086>
- Setiawan, A., Waleska, R. F., Purnama, M. A., & Efrizoni, L. (2024). Komparasi Algoritma K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), dan Decision Tree dalam Klasifikasi Penyakit Stroke. *JURNAL INFORMATIKA & REKAYASA ELEKTRONIKA*, 7(1), 107–114. <https://doi.org/10.36595/jire.v7i1.1161>

- Setiawan, R., & Triayudi, A. (2022). Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(2), 777-785. <https://doi.org/10.30865/mib.v6i2.3566>
- Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. Dalam *Machine Learning and the Internet of Medical Things in Healthcare* (hlm. 89-111). Elsevier. <https://doi.org/10.1016/B978-0-12-821229-5.00003-3>
- Supariasa, I. D. N., Bakri, B., & Fajar, I. (2016). *Penilaian Status Gizi* (2 ed.). Penerbit Buku Kedokteran ECG.
- Tarigan, A., Mustakim, Wahyudi, E., & Adhiva, J. (2019). Klasifikasi Status Kesejahteraan Rumah Tangga di Kabupaten Siak Menggunakan Algoritma Naive Bayes Classifier. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 11*. <https://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/7996>
- Triwidianti, J., Alfian, F. Y., & Prasojo, M. (2021). Perbandingan Metode Data Mining Untuk Prediksi Prestasi Siswa Tingkat Pendidikan Menengah Kejuruan Pada Sekolah Menengah Kejuruan Negeri (SMKN 1) Gadingrejo Pringsewu Lampung. *Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021, 1*, 126-133. <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2945>
- Ula, M., Ulva, A. F., Mauliza, M., Ali, M. A., & Said, Y. R. (2022). Application of Machine Learning in Determining the Classification of Children's Nutrition

- with Decision Tree. *Jurnal Teknik Informatika (Jutif)*, 3(5), 1457–1465.
<https://doi.org/10.20884/1.jutif.2022.3.5.599>
- Utomo, D. P., & Mesran. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(2), 437–444.
<https://doi.org/10.30865/mib.v4i2.2080>
- Vanacore, A., Pellegrino, M. S., & Ciardiello, A. (2024). Fair evaluation of classifier predictive performance based on binary confusion matrix. *Computational Statistics*, 39(1), 363–383. <https://doi.org/10.1007/s00180-022-01301-9>
- Yanto, M., Hadi, F., & Arlis, S. (2023). Optimization of Machine Learning Classification Analysis of Malnutrition Cases in Children. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(6), 1378–1386.
<https://doi.org/10.29207/resti.v7i6.5278>
- Yunus, M., Biddinika, M. K., & Fadlil, A. (2023). Classification of Stunting in Children Using the C4.5 Algorithm. *Jurnal Online Informatika*, 8(1), 99–106.
<https://doi.org/10.15575/join.v8i1.1062>
- Zhang, M.-L., Peña, J. M., & Robles, V. (2009). Feature selection for multi-label naive Bayes classification. *Information Sciences*, 179(19), 3218–3229.
<https://doi.org/10.1016/j.ins.2009.06.010>

LAMPIRAN

Surat permohonan penelitian ke Dinas Kesehatan Kabupaten Majalengka



UNIVERSITAS
AMIKOM
YOGYAKARTA

INSTITUSI BAHASA Indonesia
PUSAT PENELITIAN DAN PENGEMBANGAN Teknologi Informasi, Sistem Informasi, dan Sistem Komunikasi
Jalan Pahlawan (Simpang) No. 10, Yogyakarta, Indonesia
Telp. (0271) 84221106 Fax. (0271) 84221107
Email: info@amikom.ac.id www.amikom.ac.id
Pusat Penelitian dan Pengembangan Teknologi Informasi

No : 083/PASCASARJANA/AMKOM/V/2023
Lamp : -
Hal : Permohonan Penelitian

Kepada Yth:
Kepala Dinas Kesehatan Kabupaten Majalengka
Bapak H. Agus Sunanto, S.Sos., M.Si
Di tempat

Dengan Hormat,

Yang bertanda tangan dibawah ini:

Nama : Prof. Dr. Kusriani, M.Kom
Jabatan : Direktur Program Pascasarjana
NIK : 190302106

Menerangkan bahwa,

Nama : Fajar Maula Hidayat
No. Mha : 22.55.1213

Adalah Mahasiswa Program Studi Pendidikan Jarak Jauh Magister Teknik Informatika Universitas Amikom Yogyakarta yang sedang menyelesaikan Tesis, yang bersangkutan mengambil permasalahan dengan judul:

" Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Stunting Pada Balita "

Demikian surat ini kami sampaikan, dengan segala kerendahan hati kami memohon Bapak/Ibu untuk dapat memberikan ijin melaksanakan Penelitian. Atas terkabuhnya permohonan ini, kami mengucapkan terima kasih.

Yogyakarta, 8 Mei 2023

Direktur Program Pascasarjana
Universitas AMIKOM Yogyakarta



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

UNIVERSITAS AMIKOM | J. Raya Hutan Wangi, Candi Baru, Sleman,
Jawa Tengah 55221, Indonesia
Email: info@amikom.ac.id www.amikom.ac.id

Creative Economy Park

Surat permohonan penelitian ke Badan Kesatuan Bangsa dan Politik
(Bakesbangpol) Kabupaten Majalengka



UNIVERSITAS
AMIKOM
YOGYAKARTA

PRODI 8440 INFORMATIKA Teknik Informatika
PRODI 844001 INFORMATIKA Teknik Informatika, Sistem Informasi,
Kejurangan Informatika (Kejurangan), Teknik Informatika (Magister Sains),
Kejurangan Informatika (Magister dan Ilmu) (Kejurangan Informatika),
Sistem, Akuntansi, Ilmu Politik (Sistem, Ilmu Politik),
Kejurangan Informatika
PRODI 844002 INFORMATIKA Teknik Informatika, Manajemen Informatika

No : 100/PASCASARJANA/AMIKOM/2023
Lamp : -
Hal : Permohonan Penelitian

Kepada Yth:
Pimpinan Badan Kesatuan Bangsa dan Politik (BAKESBANGPOL)
Kabupaten Majalengka
Di tempat

Dengan Hormat,

Yang bertanda tangan dibawah ini:

Nama : **Prof. Dr. Kusriani, M.Kom**
Jabatan : **Direktur Program Pascasarjana**
NIK : **190302106**

Menerangkan bahwa,

Nama : **Fajar Maula Hidayat**
No. Mhs : **22.55.1213**

Adalah Mahasiswa Program Magister Teknik Informatika pada Program Studi Teknik Informatika
Program Magister Program Pendidikan Jarak Jauh Universitas AMIKOM Yogyakarta yang sedang
menyusun Tesis, yang bersangkutan mengambil permasalahan dengan judul:

" Penerapan Algoritma Naive Bayes Untuk Klasifikasi Stunting Pada Balita "

Demikian surat ini kami sampaikan, dengan segala kerendahan hati kami memohon Bapak/Ibu
untuk dapat memberikan ijin melaksanakan Penelitian. Atas terkabuleya permohonan ini, kami
mengucapkan terima kasih.

Yogyakarta, 16 Mei 2023

Direktur Program Pascasarjana
Universitas AMIKOM Yogyakarta



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

GENVA AMIKOM, Jl. Png Roudi Sora, Condong-Catur, Depok, Sleman
Telp. 0274 894211-264 Fax. 0274 894222
e-mail: info@amikom.ac.id www.amikom.ac.id
Creative Economy Park

Surat pengantar dari Bakesbangpol ke Dinas Kesehatan Kabupaten Majalengka



PEMERINTAH KABUPATEN MAJALENGKA
BADAN KESATUAN BANGSA DAN POLITIK

Jalan K.H. Abdul Halim No. 522, Majalengka 45413
Telp/Fax (0233) 282511, email : Bakesbangpol@majalengkakab.go.id
Website : bakesbangpol.majalengkakab.go.id

Majalengka, 22 Mei 2023
Kepada
Yth. Kepala Dinas Kesehatan Kab. Majalengka
di
MAJALENGKA

SURAT PENGANTAR

Nomor : PP.05.01/201/Kesbang

- A. Dasar** : 1. Peraturan Menteri Dalam Negeri Republik Indonesia Nomor 3 Tahun 2010 Tentang Penerbitan Surat Keterangan Penelitian, pasal 5 ayat (1) dan (2);
2. Surat Kepala Badan Kesatuan Bangsa dan Politik Nomor 070/310/Bakesbangpol Tanggal 27 Oktober 2021 tentang Sosialisasi Terkait Kelentuan Surat Keterangan Penelitian;
- B. Mempertimbangkan** : Surat dari Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta, Nomor 093/PASCASARJANA/AMIKOM/VI/2023, Tanggal 8 Mei 2023, Perihal Permohonan Penelitan
- Dengan ini kami sampaikan bahwa :
- | | |
|-----------------------------|---------------------|
| 1. Nama | Fajar Maula Hidayat |
| 2. Tempat, Tanggal Lahir | [REDACTED] |
| 3. Nomor Telepon | [REDACTED] |
| 4. NIM/NPM/NRP/NIDN/NPP/NPD | 22.05.1213 |
| 5. Agama | Islam |
| 6. Pekerjaan | Pelajar/Mahasiswa |
| 7. Alamat | [REDACTED] |
- E. Permintaan Data** : Penerapan agornya nawa sayes untuk kesehatan seluruh ruda Bala
- F. Lokasi** : Dinas Kesehatan Kab. Majalengka
- 10. Waktu** : 22 Mei – 31 Juli 2023
- C. Surat pengantar ini diberikan kepada yang bersangkutan, sesuai dengan ketentuan bahwa :**
- Selagi peneliti dalam melakukan penelitian harus memiliki SKP;
 - SKP dikuatkan terhadap:
 - Penelitian yang dilakukan dalam rangka tugas akhir pendidikan/ekskul di dalam negeri;
 - Penelitian yang dibuktikan instansi pemerintah yang sumber pendanaannya bersumber dari Anggaran/Pendapatan dan Belanja Negara/Anggaran Pendapatan dan Belanja Daerah;
 - Untuk penelitian sebagaimana angka 2, Pimpinan Badan Kesbangpol dan Dinas/Instansi/Badan/Lembaga/Perusahaan Daerah berhak untuk menolak apabila penelitian menimbulkan dampak negatif, serta wajib mematuhi protokol kesehatan dengan pendekatan 5 M;
 - Surat pengantar ini dibuat untuk pencatatan dan pengendalian terhadap kegiatan penelitian dan pelaksanaan pelaporan hasil kajian.
- D. Setelah melaksanakan kegiatan penelitian agar melaporkan kepada Bupati Majalengka Cq. Kepala Badan Kesatuan Bangsa dan Politik Kabupaten Majalengka berbentuk hard copy dan soft copy.**

A.n. Kepala Badan Kesatuan Bangsa dan Politik
Sekretaris Badan Kesatuan Bangsa dan Politik


Solikhudin, S.H., M.Si
Rampia
NIP. 197203261994031002

- Tersusun disampaikan kepada :
- Yth. Bupati Majalengka (sebagai laporan);
 - Yth. Kepala Dapredalibang Kabupaten Majalengka;
 - Yth. Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta.

Surat izin pendahuluan dari Dinas Kesehatan Kabupaten Majalengka



PEMERINTAH KABUPATEN MAJALENGKA DINAS KESEHATAN

Korlat - Jalan Gersam Koperasi No.44 Majalengka 45411 Telp./Faks:0223)261042
Email: dinas@pns.kabupatmajalengka.go.id

Nomor : 071/3595/SDK
Sifat : Biasa
Lampiran :
Hal : Izin Pendahuluan

Majalengka, 21 Juni 2023
Kepada :
Yth. Direktur Program
PascaSarjana Universitas
Ankren Yogyakarta
di

TEMPAT

Menyisipkan surat Saudara nomor : 093/PASCASARJANA/AMKRD/M/2023 tanggal 8 Mei 2023 perihal Permohonan Izin Pendahuluan dan Surat Pengantar Kelemb. Kerja Kesehatan Bangsa dan P469 Kabupaten Majalengka nomor : PP.05.01/201/Kesbangri tanggal 22 Mei 2023, pada prinsipnya tidak keberatan dan mengizinkan Mahasiswa/Saudara :

Nama : Fajar Maula Hilayat
NIM : 32.55.1213
Program Study : S2 Teknik Informatika Universitas Ankren Yogyakarta
Judul Penelitian : Penerapan Algoritma Naive Bayes Untuk Klasifikasi

Geometri Pada Balok
Dinas Kesehatan Kabupaten Majalengka

Lokal

Untuk melakukan pengabdian diabdikan dalam rangka penyusunan Tesis. Demikian untuk diketahui dan dapat dijadikan bahan selanjutnya.



Ditandatangani secara elektronik oleh
Kepala Dinas Kesehatan
Kabupaten Majalengka
N. ALIF 20240703.5.000.000
Pemerintah T&J

Dokumen ini telah diandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan oleh Balai Sertifikasi Elektronik (BSrE), BSSN

