

BAB III

HASIL PEMBAHASAN

3.1 Hasil Penelitian

Berdasarkan hasil penelitian yang telah dilakukan, untuk melihat pengaruh peningkatan akurasi dengan menggunakan Particle Swarm Optimization sebagai metode optimasi parameter K pada K-Nearest Neighbors di klasifikasi status gizi balita di Kota Samarinda.

3.1.1 Data Understanding

Proses data *understanding* dimulai dengan pengumpulan dan analisis data rekam medis pemeriksaan status gizi balita di Kota Samarinda. Data dikumpulkan dari Dinas Kesehatan Kota Samarinda dengan melampirkan surat pengantar dari kampus. Data ini mencakup historis pemeriksaan status gizi dari tahun 2023 hingga 2023, terdiri dari 15.593 data dalam format file excel. Terdapat 17 atribut dan 1 kelas target yaitu “BB/TB”, dimana data target yang didapat tergolong *multi-classification*.

Proses yang dilakukan peneliti selanjutnya melakukan analisis data dengan menggabungkan data menjadi satu sheet agar mudah dilakukan analisa. Langkah-langkah analisis meliputi penelitian mendalam terhadap data yang digunakan, dengan fokus fitur-fitur yang relevan untuk membangun model yang akurat. Data tahun 2023 diprioritaskan untuk keperluan analisis.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9494 entries, 0 to 9493
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   No                     9494 non-null   int64
1   Nama                   9494 non-null   object
2   JK                     9494 non-null   object
3   Kab/Kota               9494 non-null   object
4   Kec                   9494 non-null   object
5   Pukesmas               9494 non-null   object
6   Desa/Kel              9484 non-null   object
7   Posyandu               9423 non-null   object
8   Usia Saat Ukur        9494 non-null   object
9   Tanggal Pengukuran    9494 non-null   object
10  Berat                  9494 non-null   float64
11  Tinggi                 9494 non-null   float64
12  BB/U                   9494 non-null   object
13  ZS BB/U               9494 non-null   float64
14  TB/U                   9494 non-null   object
15  ZS TB/U               9494 non-null   float64
16  BB/TB                  9452 non-null   object
17  ZS BB/TB              9494 non-null   float64
18  Naik Berat Badan      9494 non-null   object
dtypes: float64(5), int64(1), object(13)
memory usage: 1.4+ MB
```

Gambar 3.1 Informasi data

Berdasarkan Gambar 3.1, menjelaskan bahwa terdapat beberapa informasi terkait isi dan type dari setiap atribut. Dengan informasi ini, peneliti dapat melanjutkan dengan tahapan mengolah data agar siap digunakan untuk membuat model. Peneliti juga menampilkan sebaran data di setiap kecamatan yang

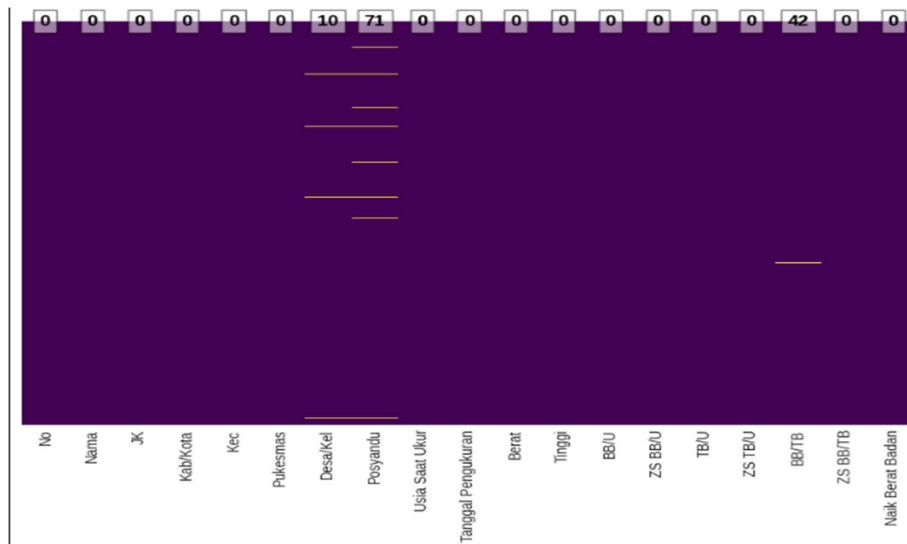
berada di Kota Samarinda berdasarkan target / kelas status gizi dari data yang diperoleh. Distribusi kelas target ditampilkan dalam Tabel 3.1 berikut:

Tabel 3.1 Informasi kelas berdasarkan kecamatan

No	Kecamatan	Gizi Baik	Gizi Buruk	Gizi Kurang	Gizi Lebih	Obesitas	Risiko Gizi Lebih	Blanks	Jumlah
1	LOA JANAN ILIR	1723	44	303	53	19	129	10	2281
2	PALARAN	665	8	87	18	14	55	1	848
3	SAMARINDA ILIR	270	14	104	19	9	38	4	458
4	SAMARINDA KOTA	270	11	53	6	4	17	1	362
5	SAMARINDA SEBERANG	565	15	84	19	7	52	8	750
6	SAMARINDA ULU	835	7	123	33	10	89	2	1099
7	SAMARINDA UTARA	494	4	46	12	5	30	3	594
8	SAMBUTAN	553	16	72	8	10	36	3	698
9	SUNGAI KUNJANG	1102	26	147	52	32	130	7	1496
10	SUNGAI PINANG	682	8	119	11	9	76	3	908
Total		7159	153	1138	231	119	652	42	9494

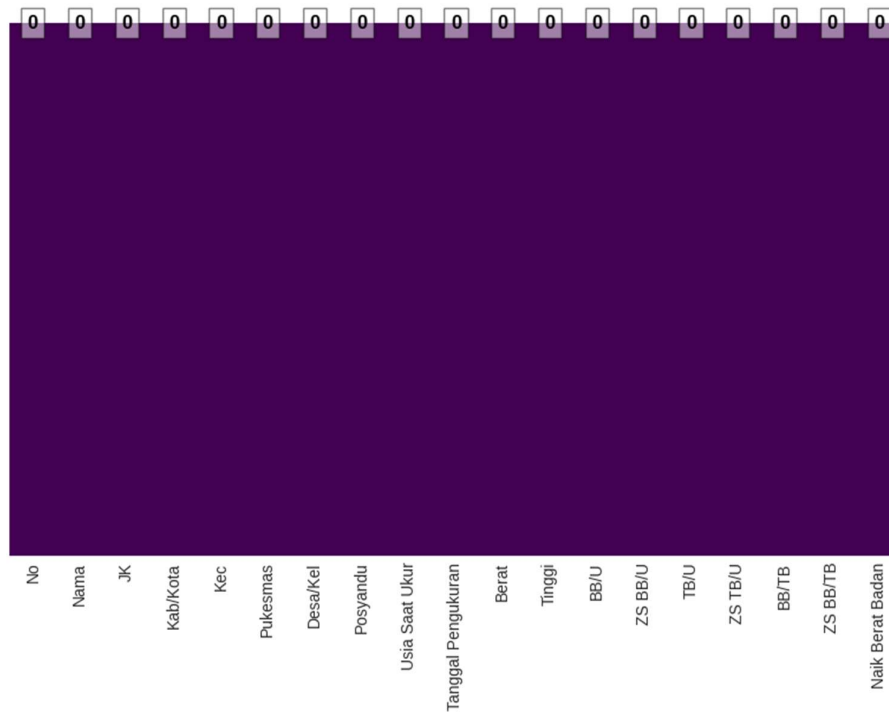
3.1.2 Data Pre-processing

Setelah mendapatkan gambaran analisis dari proses sebelumnya, peneliti melanjutkan dengan tahap *pre-processing* data. Salah satu langkah krusial dalam proses ini adalah data *cleaning*, dimana peneliti membersihkan data yang tidak diperlukan dan dapat mempengaruhi akurasi dari model yang akan dihasilkan. Langkah pertama yang dilakukan adalah pengecekan terhadap *missing value* atau nilai yang kosong dalam dataset. Data yang mengandung *missing value* dihapus agar tidak mengurangi hasil akurasi.



Gambar 3.2 Data sebelum dibersihkan

Berdasarkan Gambar 3.2, data yang mengandung *missing value* dihapus agar tidak mengurangi hasil akurasi. atribut yang memiliki *missing value* adalah “Desa/Kel”, “Posyandu” dan “BB/TB”, selanjutnya dilakukan penghapusan dengan menggunakan perintah `.dropna(inplace=True)`.



Gambar 3.3 Data setelah dibersihkan

Pada Gambar 3.3 adalah hasil data yang sudah dibersihkan, data yang mengandung *missing value* sudah tidak ada. Peneliti juga melakukan penghapusan terhadap atribut yang tidak diperlukan seperti “No”, “Nama” dan “Tanggal Pengukuran”.

Setelah melakukan proses pembersihan data dari *missing value*, maka selanjutnya proses data transformasi, terdapat beberapa atribut yang akan ditransformasi yaitu, “Kec”, “Puskesmas”, “Des/Kel”, “Posyandu”, “JK”, “BB/U”, “TB/U”, “BB/TB” dan “Naik Berat Badan”.

JK	Kec	Pukesmas	Desa/Kel	Posyandu	Usia Saat Ukur	Berat	Tinggi	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB	Naik Berat Badan	
0	P	SUNGAI KUNJANG	WONOREJO	KARANG ANYAR	HARAPAN KITA	0 Tahun - 0 Bulan - 0 Hari	3.2	45.0	Berat Badan Normal	-0.07	Pendek	-2.23	Gizi Lebih	2.77	-
1	P	SUNGAI PINANG	REMAJA	TEMINDUNG PERMAI	PULAU INDAH	4 Tahun - 0 Bulan - 16 Hari	12.0	94.0	Kurang	-2.25	Pendek	-2.09	Gizi Baik	-1.46	0
2	L	SAMBUTAN	SAMBUTAN	SAMBUTAN	AGLO NEMA	2 Tahun - 7 Bulan - 24 Hari	11.0	85.0	Berat Badan Normal	-1.81	Pendek	-2.35	Gizi Baik	-0.73	0
3	L	SAMBUTAN	SAMBUTAN	SAMBUTAN	KEMUNING	0 Tahun - 11 Bulan - 5 Hari	8.0	70.0	Berat Badan Normal	-1.52	Pendek	-2.03	Gizi Baik	-0.63	T
4	P	SAMBUTAN	SAMBUTAN	SAMBUTAN	KEMUNING	1 Tahun - 11 Bulan - 15 Hari	9.0	75.0	Berat Badan Normal	-1.96	Sangat Pendek	-3.43	Gizi Baik	-0.18	0
5	L	SAMARINDA UTARA	BENGGURUNG	SEMPAJA UTARA	AGLONEMA TERPADU	1 Tahun - 4 Bulan - 15 Hari	9.5	75.0	Berat Badan Normal	-1.01	Pendek	-2.19	Gizi Baik	0.00	0
6	L	SAMARINDA UTARA	BENGGURUNG	SEMPAJA UTARA	AGLONEMA TERPADU	2 Tahun - 0 Bulan - 20 Hari	10.4	81.3	Berat Badan Normal	-1.46	Pendek	-2.07	Gizi Baik	-0.50	0
7	P	SAMARINDA UTARA	BENGGURUNG	SEMPAJA UTARA	AGLONEMA TERPADU	2 Tahun - 10 Bulan - 29 Hari	10.5	83.7	Kurang	-2.13	Pendek	-2.82	Gizi Baik	-0.60	0
8	L	SAMARINDA UTARA	BENGGURUNG	SEMPAJA UTARA	AGLONEMA TERPADU	1 Tahun - 5 Bulan - 9 Hari	9.4	74.1	Berat Badan Normal	-1.25	Pendek	-2.81	Gizi Baik	0.11	0
9	P	SAMARINDA UTARA	BENGGURUNG	SEMPAJA UTARA	AGLONEMA TERPADU	3 Tahun - 8 Bulan - 4 Hari	11.4	91.5	Kurang	-2.32	Pendek	-2.14	Gizi Baik	-1.53	0

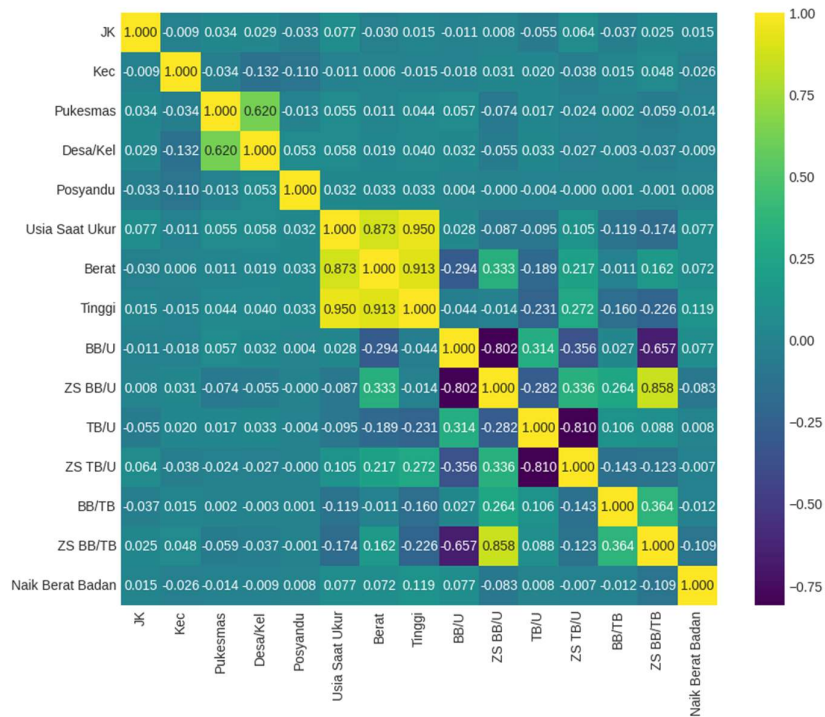
Gambar 3.4 Data sebelum ditransformasi

JK	Kec	Pukesmas	Desa/Kel	Posyandu	Usia Saat Ukur	Berat	Tinggi	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB	Naik Berat Badan	
0	1	9	26	17	168	0	3.2	45.0	0	-0.07	0	-2.23	3	2.77	0
1	1	10	16	57	347	1476	12.0	94.0	1	-2.25	0	-2.09	0	-1.46	2
2	0	8	18	35	4	964	11.0	85.0	0	-1.81	0	-2.35	0	-0.73	2
3	0	8	18	35	205	335	8.0	70.0	0	-1.52	0	-2.03	0	-0.63	3
4	1	8	18	35	205	710	9.0	75.0	0	-1.96	1	-3.43	0	-0.18	2
5	0	7	3	40	5	500	9.5	75.0	0	-1.01	0	-2.19	0	0.00	2
6	0	7	3	40	5	750	10.4	81.3	0	-1.46	0	-2.07	0	-0.50	2
7	1	7	3	40	5	1059	10.5	83.7	1	-2.13	0	-2.82	0	-0.60	2
8	0	7	3	40	5	524	9.4	74.1	0	-1.25	0	-2.81	0	0.11	2
9	1	7	3	40	5	1339	11.4	91.5	1	-2.32	0	-2.14	0	-1.53	2

Gambar 3.5 Data setelah ditransformasi

Berdasarkan Gambar 3.5, adalah proses hasil transformasi yang dilakukan pada atribut “Kec”, “Puskesmas”, “Des/Kel”, “Posyandu”, “JK”, “BB/U”, “TB/U”, “BB/TB” dan “Naik Berat Badan”, yang sebelumnya data berupa *string* diubah menjadi data berbentuk *numerik*. Hal ini bertujuan agar atribut data dapat diproses oleh algoritma KNN dan PSO untuk meningkatkan waktu dalam melakukan proses modeling data.

Selain dari atribut yang dihapus dan ditransformasi, peneliti juga melakukan perhitungan statistik dengan melihat hubungan antar atribut dengan menggunakan *correlation matrix*. Peneliti dapat melihat dan menganalisa hubungan korelasi kuat antar atribut. Dari hasil *correlation matrix*, peneliti akan menimbangakan beberapa atribut yang perlu dan tidak perlu yang digunakan untuk tahapan selanjutnya.



Gambar 3.6 Korelasi data

Berdasarkan Gambar 3.4, menjelaskan terkait hasil dari tahapan *correlation matrix* yang menampilkan hubungan antara atribut dengan kelas target. Dari hubungan ini peneliti dapat melakukan analisis singkat terkait atribut yang kurang berpengaruh terhadap kelas target. Terdapat dua hasil hubungan yang dihasilkan dari proses ini yaitu hubungan linear positif dan linear negatif. Semakin dekat dengan nilai 0 maka atribut tersebut memiliki korelasi rendah dengan kelas target. Beberapa atribut yang dihapus dalam tahapan ini adalah *Puskesmas (0,002)*, *Desa/Kel (-0,003)*, *Posyandu (0,001)*, dan *Naik Berat Badan (0,008)*.

3.1.3 Pembagian Data

Data yang sudah dilakukan proses data preprocessing selanjutnya mulai dilakukan pembagian data untuk dibagi menjadi data *training* dan data *testing*. Proses ini bertujuan untuk melakukan pengujian model terhadap data yang belum pernah dilihat sebelumnya. Rasio pembagian data yang digunakan dalam proses ini adalah sebesar 80% digunakan untuk data *training* dengan total data sebanyak 7505 dan 20% digunakan untuk data *testing* dengan total data sebanyak 1877. Pelatihan model akan diterapkan terlebih dahulu di data *training* sebelum diaplikasikan ke data *testing* maupun data baru yang ingin diuji.

3.1.4 Implementasi *K-Nearest Neighbors*

Algoritma *K-Nearest Neighbors* (KNN) merupakan kunci dalam pengembangan model untuk mengklasifikasikan status gizi balita. Dalam penelitian ini, KNN bekerja dengan cara mencari kelas mayoritas dari k tetangga terdekat dari sebuah data uji untuk menentukan label klasifikasi. Dalam mendapatkan model dengan akurasi terbaik, penting untuk melakukan pemilihan parameter yang optimal, seperti jumlah tetangga (k) dan metrik jarak yang digunakan

Untuk memperoleh konfigurasi parameter terbaik, berbagai kombinasi nilai k , *weight* (bobot), dan *distance metric* (metrik jarak) diuji. Kombinasi-kombinasi ini melibatkan nilai k yang berbeda (1 - 50), bobot ‘uniform’ dan ‘distance’, serta metrik jarak Manhattan ($p=1$) dan Euclidean ($p=2$). Tabel di bawah ini merangkum hasil pengujian berbagai kombinasi parameter tersebut, menunjukkan nilai akurasi yang diperoleh pada set pengujian (*testing set*). Dengan menganalisis tabel ini, kita dapat menentukan konfigurasi parameter KNN yang memberikan akurasi terbaik untuk tugas klasifikasi ini.

Tabel 3.2 Pengujian menggunakan KNN

No	K	Weights	Distance Metric	Testing Accuracy (%)
1	1	uniform	p=1	89,93%
2	1	uniform	p=2	89,07%
3	1	distance	p=1	89,93%
4	1	distance	p=2	89,07%
5	2	uniform	p=1	88,59%
6	2	uniform	p=2	87,63%
7	2	distance	p=1	89,93%
8	2	distance	p=2	89,07%
9	3	uniform	p=1	90,25%
10	3	uniform	p=2	88,81%
11	3	distance	p=1	90,94%
12	3	distance	p=2	89,61%
13	4	uniform	p=1	89,61%
14	4	uniform	p=2	88,54%
15	4	distance	p=1	91,15%
16	4	distance	p=2	89,93%
...
198	50	uniform	p=2	81,67%
199	50	distance	p=1	84,76%
200	50	distance	p=2	83,16%

Dari Tabel 3.2, terlihat bahwa hasil pengujian menunjukkan beberapa pola penting. Pada nilai k yang lebih rendah, model cenderung memberikan akurasi yang lebih tinggi, misalnya pada $K=3$ dengan matrik jarak Manhattan ($p=1$), akurasi mencapai 90,94%. Selain itu, penggunaan weight ‘distance’ menghasilkan akurasi yang lebih baik dibandingkan dengan ‘uniform’, contohnya pada $K=4$, ‘distance’ dengan metrik Manhattan ($p=1$) mencapai akurasi 91,15%, sementara ‘uniform’ dengan metrik yang sama hanya mencapai 89,61%.

Metrik jarak (Distance Metric) juga berpengaruh signifikan terhadap akurasi. metrik Manhattan ($p=1$) sering kali memberikan akurasi lebih tinggi dibandingkan Euclidean ($p=2$). Contohnya pada $K=4$, penggunaan ‘distance’ dengan metrik Manhattan menghasilkan akurasi 91,15%, sedangkan matrik Euclidean hanya mencapai 89,93%.

3.1.5 Implementasi *K-Nearest Neighbors & Particle Swarm Optimization*

Pada penelitian ini, algoritma KNN diimplementasikan untuk mengklasifikasikan status gizi balita, agar dapat memperoleh model KNN dengan akurasi terbaik, peneliti menggunakan metode PSO untuk mengoptimalkan parameter KNN. Parameter yang dioptimalkan meliputi jumlah tetangga

terdekat k (1 – 50), tipe *weight* dan *distance metric* (p). dalam proses optimasi, PSO menjalankan iterasi untuk mengeksplorasi ruang parameter dan menemukan kombinasi parameter yang memberikan kinerja terbaik berdasarkan *cross-validated accuracy* pada data pelatihan.

Tabel 3.3 Pengujian menggunakan KNN & PSO

No	K	Weight	Distance Metric	Mean Accuracy	Test Accuracy
1	4	Uniform	p=1	92,13%	93,71%
2	4	Distance	P=1	93,38%	94,19%
3	5	Uniform	P=1	92,47%	93,23%
4	5	Distance	P=1	93,10%	93,98%
5	6	Uniform	P=1	92,14%	92,91%
6	6	Distance	P=1	93,44%	93,98%
7	7	Uniform	P=1	92,35%	92,81%
8	7	Distance	P=1	93,10%	93,77%
9	8	Uniform	P=1	91,75%	92,75%
10	8	Distance	P=1	93,16%	93,93%

Pada Tabel 3.3, menunjukkan hasil optimasi dengan parameter optimal adalah $k=6$, *weight* 'distance', dan $p=1$. Dengan parameter ini, model KNN mencapai *mean Accuracy* sebesar 93,44%. Setelah model KNN dengan parameter optimal ini di-fit pada data pelatihan, model tersebut diuji pada data uji untuk mengukur performa sesungguhnya. Akurasi pada data *training* adalah 100%, sedangkan akurasi pada data *testing* adalah 93,98%. Nilai akurasi pengujian yang tinggi menunjukkan bahwa model KNN yang dioptimalkan dengan PSO dapat menggeneralisasi dengan baik terhadap data yang belum pernah dilihat sebelumnya.

Selain itu, selama proses optimasi, hasil dari berbagai kombinasi parameter juga dicatat. Misalnya, iterasi dengan $k=4$, *weight* 'distance', dan $p=1$ menghasilkan *mean accuracy* sebesar 93,38%, sedangkan iterasi dengan $k=8$, *weight* 'distance', dan $p=1$ menghasilkan *mean accuracy* sebesar 93,16%. Analisis dari berbagai iterasi ini memberikan gambaran bahwa parameter k yang lebih rendah cenderung memberikan hasil yang lebih baik ketika menggunakan *weight* 'distance'.

3.1.6 Evaluasi Model

Tahapan selanjutnya adalah melakukan evaluasi pada model KNN yang diterapkan untuk mengklasifikasi status gizi balita menunjukkan bahwa pemilihan parameter yang tepat sangat krusial dalam meningkatkan akurasi model. Berdasarkan hasil pengujian berbagai kombinasi parameter KNN, ditemukan bahwa kombinasi terbaik adalah $K=4$, bobot 'distance', dan metrik jarak $p=1$, yang memberikan akurasi tertinggi pada data uji sebesar 91,15%. Tren dari hasil pengujian ini menunjukkan bahwa penggunaan bobot 'distance' secara konsisten menghasilkan performa yang lebih baik dibandingkan dengan bobot 'uniform', dan metrik jarak $p=1$ cenderung lebih akurat daripada metrik jarak $p=2$ untuk sebagian besar nilai k yang diuji.

Implementasi KNN dengan optimasi menggunakan Particle Swarm Optimization (PSO) menghasilkan parameter optimal dengan $k=6$, *weight* 'distance', dan $p=1$, yang memberikan *mean accuracy* 93,44% dan *test accuracy* 93,98%. Hasil optimasi ini menunjukkan kemampuan generalisasi yang baik terhadap data uji. Iterasi lain dengan $k=4$ dan $k=8$, juga menunjukkan performa yang sangat baik, namun nilai $k=6$ terbukti memberikan kombinasi terbaik. PSO terbukti efektif dalam mengeksplorasi ruang parameter dan menemukan kombinasi yang memberikan performa optimal.

3.1.7 Pengujian Rasio

Tahapan selanjutnya dengan melakukan pengujian terhadap rasio pembagian data *training* dan data *testing*. Tahap ini dilakukan untuk mengetahui rasio pembagian data terbaik terhadap model klasifikasi dan dataset yang digunakan. Pada pengujian rasio data total data yang digunakan sebanyak 9.382 data yang didapat dari hasil pengolahan data *preprocessing* yang sudah dilakukan pada tahap sebelumnya. Akurasi terbaik dari hasil pengujian sebesar 93,98%, dengan rasio pembagian data 80% : 20% dimana data *training* berjumlah 7.505 data dan data *testing* berjumlah 1.877 data. Hasil dari pengujian rasio data ditampilkan dalam Tabel 3.4, sebagai berikut:

Tabel 3.4 Pengujian Rasio Data

Pengujian Rasio Data	Jumlah Data Training	Jumlah Data Testing	Akurasi
50% : 50%	4691	4691	92,84%
60% : 40%	5629	3753	92,94%
70% : 30%	6567	2851	92,93%
80% : 20%	7505	1877	93,98%
90% : 10%	8443	939	93,82%

3.1.8 Hasil dan Validasi

Hasil dan validasi merupakan tahap akhir dalam penelitian ini, yang dilakukan dengan menganalisis hasil dari *confusion matrix* untuk menentukan nilai TP (*True Positive*), FP (*False Negative*), FN (*False Negative*), dan TN (*True Negative*).

Gambar 3.7 *Confusion Matrix*



Berdasarkan Gambar 3.5, nilai dari *confusion matrix* selanjutnya divalidasi kebenarannya untuk memastikan apakah hasil akurasi yang diperoleh sama dengan hasil perhitungan yang dilakukan menggunakan tabel *confusion matrix*. Perhitungan akurasi dilakukan dengan menjumlahkan semua prediksi yang benar (*True Positive*) dan membagi dengan total data. Perhitungan akurasi dari hasil *confusion matrix* adalah sebagai berikut :

$$Accuracy = \frac{Correct\ Prediction}{Total\ Instance}$$

$$Accuracy = \frac{(1421 + 18 + 187 + 32 + 21 + 85)}{1877}$$

$$Accuracy = \frac{1.764}{1877}$$

$$Accuracy = 0.9398 = 93.98\%$$

Hasil perhitungan akurasi kemudia divalidasi kebenarannya terhadap akurasi terhadap model klasifikasi. Hal ini untuk memastikan bahwa akurasi memang sudah sesuai perhitungannya dan tidak ada manipulasi data didalamnya, sehingga penelitian ini dapat dianggap.

3.1.9 Perbandingan Hasil

Dalam melakukan perbandingan antara dua model *K-Nearest Neighbors*, di mana satu menggunakan optimasi *Particle Swarm Optimizaiton* (PSO) dan satunya lagi tidak, fokus pada perbedaan akurasi memberikan pandangan yang penting dalam mengevaluasi efektivitas optimasi dalam meningkatkan kinerja model. Sehingga menghasilkan akurasi yang lebih akurat dan efisien

Tabel 3.5 Perbandingan Hasil akurasi pengujian KNN

K	KNN	KNN+PSO	Status
4	91,15%	94,19%	Naik
5	90,88%	93,98%	Naik
6	90,88%	93,98%	Naik
7	90,72%	93,77%	Naik
8	90,41%	93,93%	Naik
9	90,03%	93,66%	Naik
10	89,98%	94,14%	Naik

Pada Tabel 3.5 menampilkan hasil perbandingan akurasi pengujian antara metode *K-Nearest Neighbors* (KNN) dan dengan optimasi *Particle Swarm Optimization* (PSO), terlihat bahwa metode dengan optimasi cenderung memberikan hasil akurasi yang lebih tinggi.

3.2 Pembahasan

Penelitian ini menunjukkan bahwa optimasi algoritma *K-Nearest Neighbors* (KNN) menggunakan *Particle Swarm Optimization* (PSO) menghasilkan peningkatan akurasi yang signifikan dalam klasifikasi status gizi balita di Kota Samarinda. Dengan parameter $K = 6$, *weight 'distance*, dan metrik jarak $P = 1$, diperoleh akurasi sebesar 93,98%, ini jauh lebih tinggi dibandingkan dengan penelitian yang dilakukan oleh Setiawan et al (2019) yang menggunakan *Naïve Bayes* dengan 412 data, yang mencapai akurasi 80,60%. Selain itu penelitian lain oleh Fitrianiingsih et al (2021) yang menggunakan KNN tanpa optimasi hanya mencapai akurasi sebesar 73,53% pada $K=3$ dan $K=5$.

Perbandingan ini menunjukkan bahwa optimasi KNN dengan PSO secara signifikan meningkatkan performa model dalam klasifikasi status gizi balita dibandingkan dengan metode sebelumnya. Hasil ini menunjukkan efektivitas PSO dalam menemukan konfigurasi parameter yang optimal dan memberikan akurasi yang lebih tinggi.