

# **Laporan Statistika Inferensi**

## **Project Akhir**



### **Dosen Pengampu:**

Dr. Candra Dewi, S.Kom., M.Sc.

### **Kelompok:**

Haikal Thoriq Athaya	225150200111004
Gede Indra Adi Brata	225150200111006
Javier Aahmes Reansyah	225150201111001
Muhammad Herdi Adam	225150201111003

**Program Studi Teknik Informatika**

**Fakultas Ilmu Komputer**

**Tahun 2024/2025**

# Daftar Isi

Daftar Isi.....	1
Deskripsi Project Laporan 1:.....	2
Sampling dan Distribusi Data.....	3
Pengecekan Normalitas Metode Kolmogorov-Smirnov.....	6
Transformasi Data.....	14
Visualisasi Setelah Normalisasi.....	19

## Deskripsi Project Laporan 1:

1. Setiap kelompok terdiri dari 3-4 mahasiswa (maksimal ada 8 kelompok) dengan ketentuan dataset adalah sebagai berikut:
  - a. Kelompok 1-3 : dataset project1
  - b. Kelompok 4-6 : dataset project2
  - c. Kelompok 7-8 : dataset project3
2. Lakukan sampling untuk menentukan data latih dan uji (misal dengan perbandingan 80:20) dan lakukan pengecekan distribusi pada data sedemikian hingga sample data latih memiliki distribusi yang sama (atau mendekati) dengan keseluruhan data.
3. Lakukan pengecekan normalitas menggunakan metode Kolmogorov-Smirnov terhadap data untuk memastikan normalitas data (terutama data latih).
4. Berdasarkan distribusi dan normalitas datanya, lakukan transformasi yang sesuai untuk data/fitur yang tidak distribusinya tidak normal.

# Sampling dan Distribusi Data

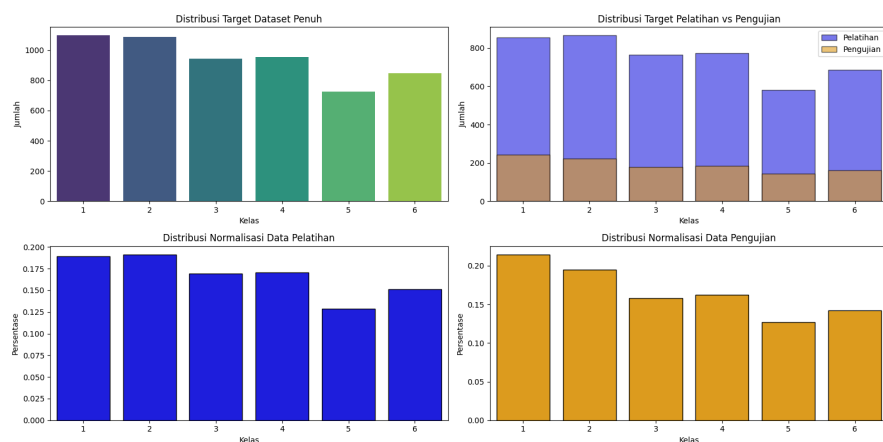
## Soal:

Lakukan sampling untuk menentukan data latih dan uji (misal dengan perbandingan 80:20) dan lakukan pengecekan distribusi pada data sedemikian hingga sample data latih memiliki distribusi yang sama (atau mendekati) dengan keseluruhan data.

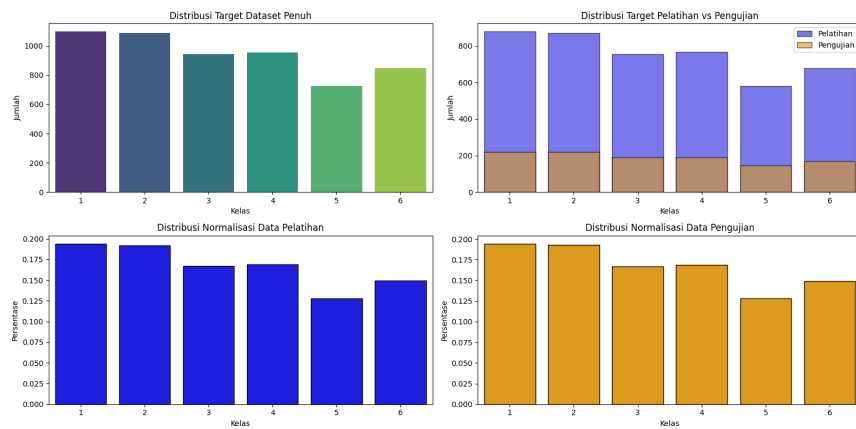
## Penjelasan:

Sampling adalah proses pengambilan sebagian data dari suatu populasi atau dataset untuk dianalisis. Tujuan dari sampling adalah mendapatkan gambaran yang representatif dari keseluruhan data tanpa harus menganalisis seluruh dataset. Distribusi data adalah bagaimana data tersebar atau didistribusikan dalam suatu dataset. Distribusi ini, dapat membantu peneliti memahami pola, karakteristik dari data yang diteliti.

Untuk memastikan bahwa data latih dan data uji memiliki distribusi kelas yang seimbang, dilakukan proses pembagian data dengan rasio 80:20 menggunakan teknik stratifikasi. Stratifikasi ini diterapkan untuk membagi data sedemikian rupa sehingga setiap kelas dalam data latih dan data uji memiliki proporsi yang mirip dengan keseluruhan data, menjaga agar distribusi kelas tidak terlalu berbeda antara subset.



**Gambar 1.1 Distribusi Kelas Tanpa Stratifikasi**



**Gambar 1.2 Distribusi Kelas Dengan Stratifikasi**

**Tabel 1.1 Distribusi Kelas Dengan Stratifikasi dan Tanpa Stratifikasi**

Kelas	Distribusi Kelas Tanpa Stratifikasi		Distribusi Kelas Dengan Stratifikasi	
	Data Latih	Data Uji	Data Latih	Data Uji
1	0.188992	0.214664	0.194076	0.194346
2	0.191424	0.195230	0.192087	0.192580
3	0.169098	0.158127	0.166888	0.166961
4	0.170645	0.162544	0.169098	0.168728
5	0.128426	0.127208	0.128205	0.128092
6	0.151415	0.142226	0.149646	0.149293

Hasil dari proses sampling menunjukkan bahwa tanpa stratifikasi, distribusi kelas pada data latih dan data uji cenderung bervariasi secara signifikan. Sebagai contoh, pada kelas 1, proporsi kelas di data latih adalah 18.90%, sedangkan pada data uji menjadi 21.47%. Namun, dengan stratifikasi, distribusi kelas pada data latih dan data uji menjadi lebih konsisten dan mendekati distribusi data keseluruhan.

Misalnya, untuk kelas 1 setelah stratifikasi, distribusi pada data latih adalah 19.41% dan pada data uji 19.43%, yang hampir sama dengan distribusi pada data awal.

Dengan demikian, teknik stratifikasi ini memastikan bahwa perbedaan proporsi kelas antara data latih dan data uji diminimalkan, sehingga model dapat dilatih pada data yang lebih representatif dan terhindar dari bias kelas.

# Pengecekan Normalitas Metode Kolmogorov-Smirnov

## Soal:

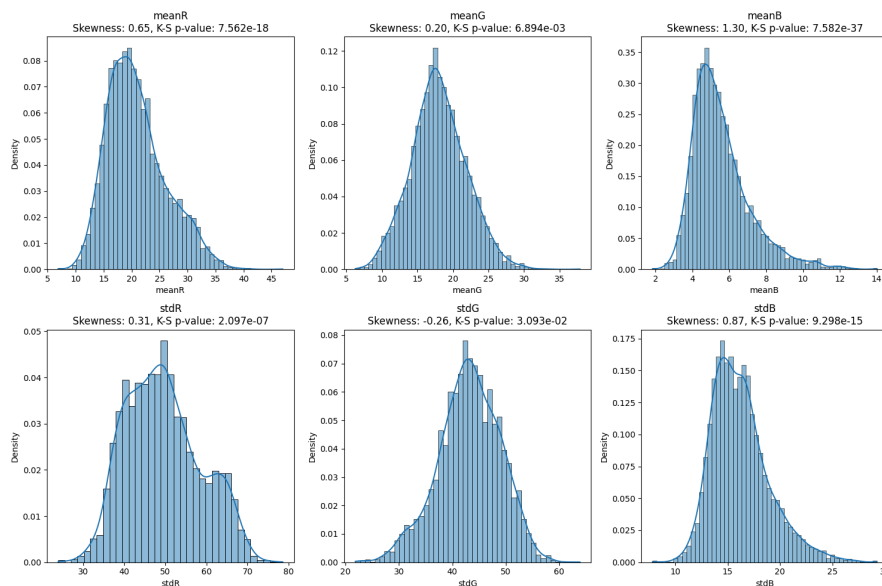
Lakukan pengecekan normalitas menggunakan metode *Kolmogorov-Smirnov* terhadap data untuk memastikan normalitas data (terutama data latih).

## Penjelasan:

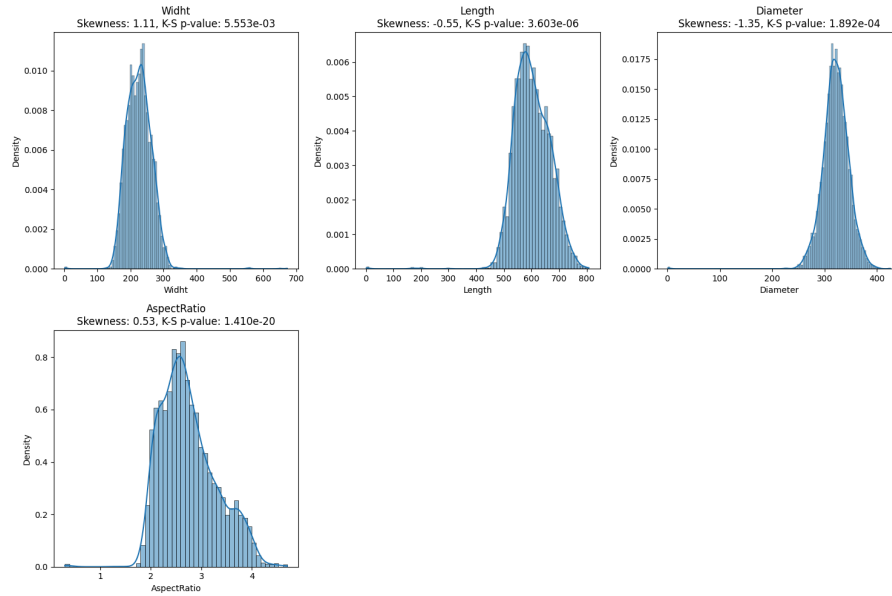
Normalitas adalah kondisi yang di mana suatu data mengikuti pola distribusi normal. Distribusi normal adalah distribusi probabilitas yang simetris dan berbentuk *bell curve*. Metode *Kolmogorov-Smirnov* adalah teknik yang digunakan dalam mengecek normalitas data. Uji ini membandingkan distribusi kumulatif dari dataset dengan distribusi teoritis untuk menentukan seberapa besar perbedaan.

a. Dibawah ini adalah gambar distribusi data yang sebelum dilakukan normalisasi

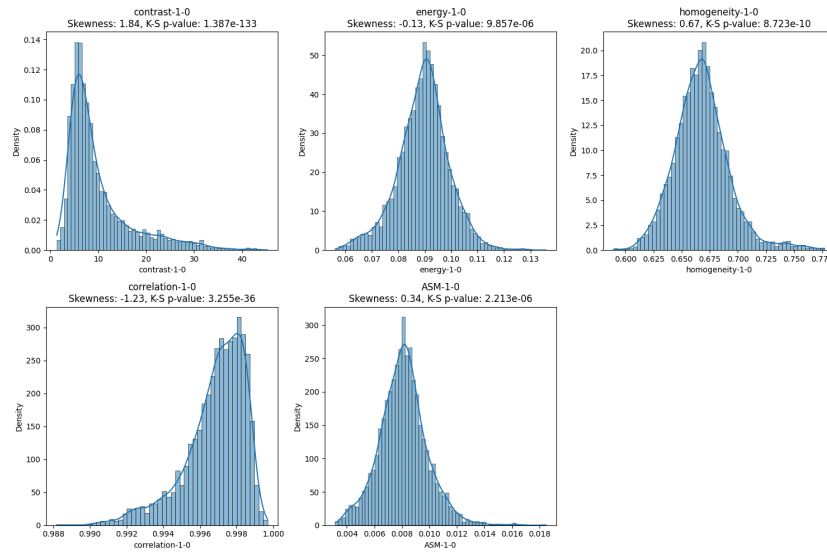
- [https://drive.google.com/file/d/1-ZvzWiW466B\\_z9\\_IvX6VQqRcS4L-nFwh/view?usp=sharing](https://drive.google.com/file/d/1-ZvzWiW466B_z9_IvX6VQqRcS4L-nFwh/view?usp=sharing) (sebelum dilakukan normalisasi)
- [https://drive.google.com/file/d/16zRwhj00ycdezVKZmJ2bnU5SB\\_YrFOkR/view?usp=drive\\_link](https://drive.google.com/file/d/16zRwhj00ycdezVKZmJ2bnU5SB_YrFOkR/view?usp=drive_link) (sesudah dilakukan normalisasi)



**Gambar 2.1 Histogram Fitur Statistik Warna pada Data Latih Mentah**

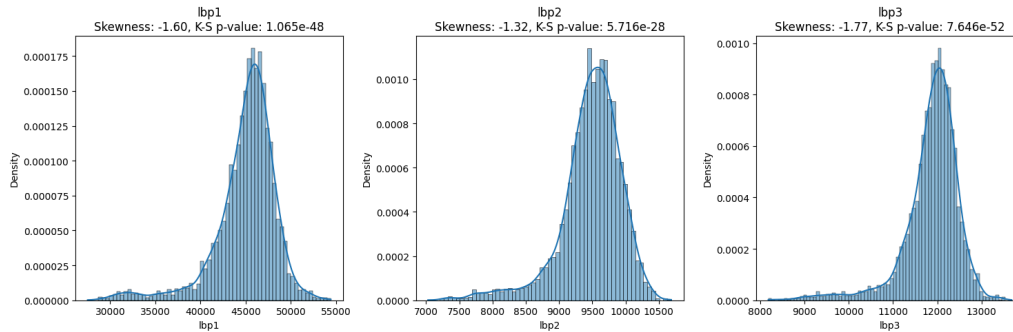


**Gambar 2.2 Histogram Fitur Bentuk dan Geometri pada Data Latih Mentah**





**Gambar 2.3 Histogram Fitur GLCM pada Data Latih Mentah**



**Gambar 2.4 Histogram Fitur LBP pada Data Latih Mentah**

**2.1 Tabel Nilai Kolmogorov Smirnov Pada Data Latih Mentah**

	<b>K-S Statistic</b>	<b>p-value</b>	<b>Skewness</b>
meanR	0.066518404	7.56E-18	0.654869386
meanG	0.024995861	0.006893651	0.203039581
meanB	0.096140253	7.58E-37	1.302139207
stdR	0.042100291	2.10E-07	0.314865196
stdG	0.021428827	0.030926033	-0.25695512
stdB	0.060334264	9.30E-15	0.869692555
skewR	0.035786743	1.81E-05	0.537395574
skewG	0.067239242	3.15E-18	0.928699206

skewB	0.018188001	0.099034972	0.147743225
kurR	0.064280015	1.08E-16	1.130779676
kurG	0.109326133	1.51E-47	1.783804756
kurB	0.034058845	5.39E-05	0.957581439
entR	0.02804488	0.001591882	-0.289534341
entG	0.02594332	0.004451239	-0.305356895
entB	0.04258733	1.44E-07	0.340089352
Widht	0.02546853	0.005553417	1.107421313
Length	0.038192228	3.60E-06	-0.546693471
Diameter	0.031961239	0.000189233	-1.349490555
Perimeter	0.032183924	0.000166257	-1.135190416
Area	0.023977762	0.01083207	0.052947815
AspectRatio	0.071536524	1.41E-20	0.525453923
FormFactor	0.0161513	0.186605743	0.695298771
Rectangularity	0.039657068	1.28E-06	3.46170573
Narrow Factor	0.09611787	7.88E-37	11.86541641
RatioOfDiameter	0.036350199	1.25E-05	0.210912875
RatioPLPW	0.045922167	9.94E-09	-4.040842445
Solidity	0.0706625	4.35E-20	-0.718809568
Convexity	0.496269318	0	46.49732433

lbp1	0.1106518	1.07E-48	-1.595098031
lbp2	0.08362296	5.72E-28	-1.316614237
lbp3	0.114189254	7.65E-52	-1.772972333
lbp4	0.043836252	5.43E-08	-0.215345438
lbp5	0.060718145	6.10E-15	-0.161079374
lbp6	0.050617968	1.63E-10	0.441771153
lbp7	0.065327824	3.14E-17	0.979682015
lbp8	0.046976388	4.09E-09	0.456903187
lbp9	0.075685596	5.52E-23	1.090903301
lbp10	0.04577498	1.12E-08	0.552992977
lbp11	0.061851453	1.73E-15	0.895401679
lbp12	0.077535668	4.21E-24	1.202437126
lbp13	0.061180513	3.66E-15	1.071174154
lbp14	0.117711756	4.50E-55	1.998793651
lbp15	0.048055978	1.62E-09	0.750823808
lbp16	0.111905674	8.40E-50	1.745982369
lbp17	0.055724484	1.19E-12	0.613412041
lbp18	0.102178915	1.42E-41	1.619951944
lbp19	0.080341469	7.56E-26	1.092935299
lbp20	0.075278253	9.65E-23	1.108475662

lbp21	0.087947515	6.77E-31	1.40159646
lbp22	0.015734946	0.210408317	0.001962857
lbp23	0.05018829	2.42E-10	0.397514659
lbp24	0.097924236	3.26E-38	-1.450992804
lbp25	0.025230809	0.006194407	-0.169901384
lbp26	0.081567954	1.25E-26	-1.156155653
contrast-1-0	0.183355518	1.39E-133	1.836046734
energy-1-0	0.036709422	9.86E-06	-0.131421932
homogeneity-1-0	0.048758017	8.72E-10	0.671373972
correlation-1-0	0.09530291	3.25E-36	-1.228433044
ASM-1-0	0.038890089	2.21E-06	0.342531765
dissimilarity-1-0	0.149820908	4.11E-89	1.294565162
contrast-1-45	0.163687455	2.17E-106	1.602992202
energy-1-45	0.039124543	1.87E-06	-0.232915902
homogeneity-1-45	0.044200567	4.06E-08	0.540399205
correlation-1-45	0.071809329	9.89E-21	-0.913833533
ASM-1-45	0.0354715	2.22E-05	0.227031953
dissimilarity-1-45	0.140690698	1.37E-78	1.279168959
contrast-1-90	0.165124738	2.86E-108	1.570736741
energy-1-90	0.040301388	8.04E-07	-0.3078971

homogeneity-1-90	0.034447896	4.24E-05	0.222679343
correlation-1-90	0.07557642	6.41E-23	-0.917263461
ASM-1-90	0.02259637	0.019407639	0.118750415
dissimilarity-1-90	0.138039338	1.17E-75	1.257207898
contrast-1-135	0.164159428	5.27E-107	1.620320873
energy-1-135	0.036687368	1.00E-05	-0.219863471
homogeneity-1-135	0.048413742	1.18E-09	0.624777019
correlation-1-135	0.079379093	3.05E-25	-0.953420912
ASM-1-135	0.032027162	0.000182136	0.2361375
dissimilarity-1-135	0.140972915	6.65E-79	1.275603656

Hasil uji *Kolmogorov-Smirnov* (K-S) yang dilakukan terhadap berbagai fitur menunjukkan perbedaan signifikan dalam distribusi data masing-masing fitur. *P-value* yang sangat kecil (misalnya, 7.56E-18 untuk meanR) menunjukkan bahwa sebagian besar distribusi data pada fitur tersebut tidak mengikuti distribusi normal, karena *p-value* yang lebih kecil dari 0.05 mengindikasikan penolakan terhadap hipotesis nol, yaitu bahwa data mengikuti distribusi normal. Beberapa fitur seperti meanG, meanB, dan stdG menunjukkan *p-value* yang lebih besar dari 0.05, yang menunjukkan bahwa distribusi data pada fitur-fitur ini lebih mendekati distribusi normal. Selain itu, nilai *skewness* pada fitur-fitur ini memberikan gambaran mengenai simetri distribusi data, di mana sebagian besar fitur menunjukkan *skewness* positif, menunjukkan distribusi data yang cenderung lebih terpusat pada sisi kiri dan memiliki ekor yang lebih panjang di sisi kanan. Beberapa fitur, seperti AspectRatio dan Narrow Factor, memiliki *skewness* yang sangat tinggi, mengindikasikan distribusi data yang sangat tidak simetris. Nilai kurtosis yang lebih besar dari 3 pada beberapa fitur, seperti Rectangularity dan Narrow Factor, menunjukkan bahwa distribusi data memiliki puncak yang lebih tajam dibandingkan distribusi normal. Secara keseluruhan, hasil ini memberikan gambaran yang lebih jelas

mengenai karakteristik distribusi data dan membantu dalam menentukan teknik analisis atau transformasi data yang tepat untuk model selanjutnya.

# Transformasi Data

## Soal:

Berdasarkan distribusi dan normalitas datanya, lakukan transformasi yang sesuai untuk data/fitur yang tidak distribusinya tidak normal.

## Penjelasan:

Kelompok kami menggunakan 4 metode yaitu *Box-Cox*, *log*, *square*, dan *Yeo-Johnson* transformasi untuk memastikan bahwa distribusi data menjadi lebih normal.

*Box-Cox* Transformasi adalah teknik yang digunakan dalam mengubah data agar distribusinya mendekati distribusi normal. Metode ini berguna saat mempunyai data positif dan tidak mengikuti distribusi normal. Cara kerja *Box-Cox* yaitu dengan mengubah nilai-nilai dalam dataset menggunakan fungsi yang tergantung pada parameter tertentu.

*Log* Transformasi adalah metode yang mengubah data dengan mengambil logaritma dari setiap nilai dalam dataset. Cara kerja *Log* yaitu pada data yang digunakan memiliki distribusi *skewed* atau memiliki rentang yang sangat luas.

*Square* Transformasi adalah metode yang mengubah data dengan mengkuadratkan setiap nilai dalam dataset. Transformasi *Square* ini dapat meningkatkan variabilitas data. Cara kerja *Square* Transformasi ini digunakan saat memperbesar perbedaan antar nilai dalam data, terutama data yang memiliki nilai-nilai yang sangat kecil atau meningkatkan distribusi data yang tidak normal.

*Yeo-Johnson* Transformasi adalah metode yang digunakan pada data yang memiliki nilai nol atau negatif, yang membuat *Yeo-Johnson* Transformasi lebih fleksibel dan dapat digunakan dalam berbagai situasi. Cara kerja dari *Yeo-Johnson* Transformasi ini yaitu dengan mengubah nilai-nilai dalam dataset sehingga distribusinya menjadi lebih mendekati distribusi normal. Dan dengan menggunakan *Box-Cos*, *log*, *square*, dan *Yeo-Johnson* transformasi, akan dapat mengoptimalkan distribusi data.

**Tabel 3.1 Fitur yang Berhasil dan Gagal Dinormalisasi**

<b>Metode Transformasi</b>	<b>Fitur Yang Sukses Dinormalisasi</b>	<b>Fitur Yang Gagal Dinormalisasi</b>
Box Cox	meanG, meanB, stdG, stdB, skewR, skewG, entR, entG, Widht, Perimeter, lbp22	meanR, stdR, skewB, kurB, Length, Diameter, Area, Rectangularity, RatioOfDiameter, lbp1, lbp2, lbp3, lbp4, lbp5, lbp6, lbp7, lbp8, lbp9, lbp10, lbp12, lbp13, lbp14, lbp15, lbp16, lbp17, lbp18, lbp20, lbp23, lbp24, lbp25, lbp26, contrast-1-0, contrast-1-45, contrast-1-90, contrast-1-135
Log Test	meanR, skewR, lbp22	meanG, meanB, stdR, stdG, stdB, skewG, skewB, kurB, entR, entG, entB, Widht, Length, Diameter, Perimeter, Area, AspectRatio, FormFactor, Rectangularity, Narrow Factor, RatioOfDiameter, RatioPLPW, Solidity, Convexity, lbp1, lbp2, lbp3, lbp4, lbp5, lbp6, lbp7, lbp8, lbp9, lbp10, lbp11, lbp12, lbp13, lbp14, lbp15, lbp16, lbp17, lbp18, lbp19, lbp20, lbp21, lbp23, lbp24, lbp25, lbp26, contrast-1-0, energy-1-0, homogeneity-1-0, correlation-1-0, ASM-1-0, dissimilarity-1-0, contrast-1-45, energy-1-45, homogeneity-1-45, correlation-1-45, ASM-1-45, dissimilarity-1-45, contrast-1-90, energy-1-90, homogeneity-1-90, correlation-1-90, ASM-1-90, dissimilarity-1-90, contrast-1-135, energy-1-135, homogeneity-1-135, correlation-1-135, ASM-1-135, dissimilarity-1-135
Square Test	kurB, Widht, lbp22	meanR, meanG, meanB, stdR, stdG, stdB, skewR, skewG, skewB, entR, entG, entB, Length, Diameter, Perimeter, Area, AspectRatio, FormFactor, Rectangularity, Narrow Factor, RatioOfDiameter, RatioPLPW, Solidity, Convexity, lbp1, lbp2, lbp3, lbp4, lbp5, lbp6, lbp7, lbp8, lbp9, lbp10, lbp11, lbp12, lbp13, lbp14, lbp15, lbp16, lbp17, lbp18, lbp19, lbp20, lbp21, lbp23, lbp24, lbp25, lbp26, contrast-1-0, energy-1-0,



		homogeneity-1-0, correlation-1-0, ASM-1-0, dissimilarity-1-0, contrast-1-45, energy-1-45, homogeneity-1-45, correlation-1-45, ASM-1-45, dissimilarity-1-45, contrast-1-90, energy-1-90, homogeneity-1-90, correlation-1-90, ASM-1-90, dissimilarity-1-90, contrast-1-135, energy-1-135, homogeneity-1-135, correlation-1-135, ASM-1-135, dissimilarity-1-135
Yeo Johnson	meanG, meanB, stdG, stdB, skewR, skewG, skewB, kurR, kurG, kurB, entR, entG, Diameter, Perimeter, Area, FormFactor, RatioOfDiameter, RatioPLPW, lbp10, lbp17, lbp22, energy-1-90, ASM-1-90	meanR, stdR, entB, Widht, Length, AspectRatio, Rectangularity, Narrow Factor, Solidity, Convexity, lbp1, lbp2, lbp3, lbp4, lbp5, lbp6, lbp7, lbp8, lbp9, lbp11, lbp12, lbp13, lbp14, lbp15, lbp16, lbp18, lbp19, lbp20, lbp21, lbp23, lbp24, lbp25, lbp26, contrast-1-0, energy-1-0, homogeneity-1-0, correlation-1-0, ASM-1-0, dissimilarity-1-0, contrast-1-45, energy-1-45, homogeneity-1-45, correlation-1-45, ASM-1-45, dissimilarity-1-45, contrast-1-90, homogeneity-1-90, correlation-1-90, dissimilarity-1-90, contrast-1-135, energy-1-135, homogeneity-1-135, correlation-1-135, ASM-1-135, dissimilarity-1-135
Kombinasi	meanG, meanB, stdG, stdB, skewR, skewG, entR, entG, Widht, Perimeter, lbp22, correlation-1-90, correlation-1-135	meanR, stdR, skewB, kurR, kurG, kurB, entB, Length, Diameter, Area, AspectRatio, FormFactor, Rectangularity, Narrow Factor, RatioOfDiameter, RatioPLPW, Solidity, Convexity, lbp1, lbp2, lbp3, lbp4, lbp5, lbp6, lbp7, lbp8, lbp9, lbp10, lbp11, lbp12, lbp13, lbp14, lbp15, lbp16, lbp17, lbp18, lbp19, lbp20, lbp23, lbp24, lbp25, lbp26, contrast-1-0, energy-1-0, homogeneity-1-0, correlation-1-0, ASM-1-0, dissimilarity-1-0, contrast-1-45, energy-1-45, homogeneity-1-45, correlation-1-45, ASM-1-45, dissimilarity-1-45, contrast-1-90, energy-1-90, homogeneity-1-90, ASM-1-90, dissimilarity-1-90, contrast-1-135, energy-1-135, homogeneity-1-135, ASM-1-135, dissimilarity-1-135

Dalam analisis normalisasi fitur menggunakan 4 metode transformasi (*Box-Cox*, *Log Test*, *Square Test*, *Yeo-Johnson*), hasil yang didapatkan menunjukkan perbedaan dalam kemampuan metode untuk menormalkan fitur. Penentuan keberhasilan normalisasi didasarkan pada dua kriteria: *p-value* dari uji *Kolmogorov-Smirnov* (K-S) yang lebih besar dari *threshold*  $\alpha$  (0.05), dan *skewness* yang berada dalam kisaran -1 hingga 1.

*Box-Cox* Transformasi cenderung efektif dalam menormalkan fitur dengan distribusi yang lebih mirip dengan distribusi normal, seperti meanG, meanB, dan skewR, di mana nilai *p-value* uji K-S lebih besar dari 0.05 dan *skewness* terkontrol. Namun, fitur seperti meanR, stdR, dan skewB gagal dinormalisasi karena distribusinya terlalu jauh dari distribusi normal, atau nilai *skewness*nya lebih besar dari 1, yang mengindikasikan ketidaksesuaian distribusi tersebut dengan normalitas yang diharapkan. *Box-Cox* efektif untuk fitur dengan distribusi yang bisa ditransformasi lebih dekat ke bentuk normal, tetapi tidak cukup baik untuk distribusi dengan *skewness* tinggi atau distribusi dengan pencilan yang signifikan.

*Log Test* menunjukkan hasil yang lebih baik pada fitur seperti meanR dan skewR yang berhasil dinormalisasi, namun gagal pada fitur lain yang memiliki nilai *skewness* tinggi atau distribusi yang tidak dapat disesuaikan dengan transformasi *log*. Metode ini sering kali lebih efektif pada data yang memiliki nilai kecil dan positif karena logaritma tidak dapat diterapkan pada angka negatif atau nol, yang membatasi kemampuannya dalam menangani berbagai jenis distribusi.

*Square Test*, yang menggunakan akar kuadrat, memiliki hasil yang lebih terbatas, dengan beberapa fitur seperti kurB dan Widht gagal dinormalisasi. Transformasi akar kuadrat dapat bekerja dengan baik pada distribusi dengan *skewness* sedang atau sedikit miring, tetapi tidak efektif untuk distribusi dengan pencilan atau distribusi yang sangat miring. Penggunaan metode ini sering kali tergantung pada seberapa dekat distribusi awal dengan bentuk distribusi normal.

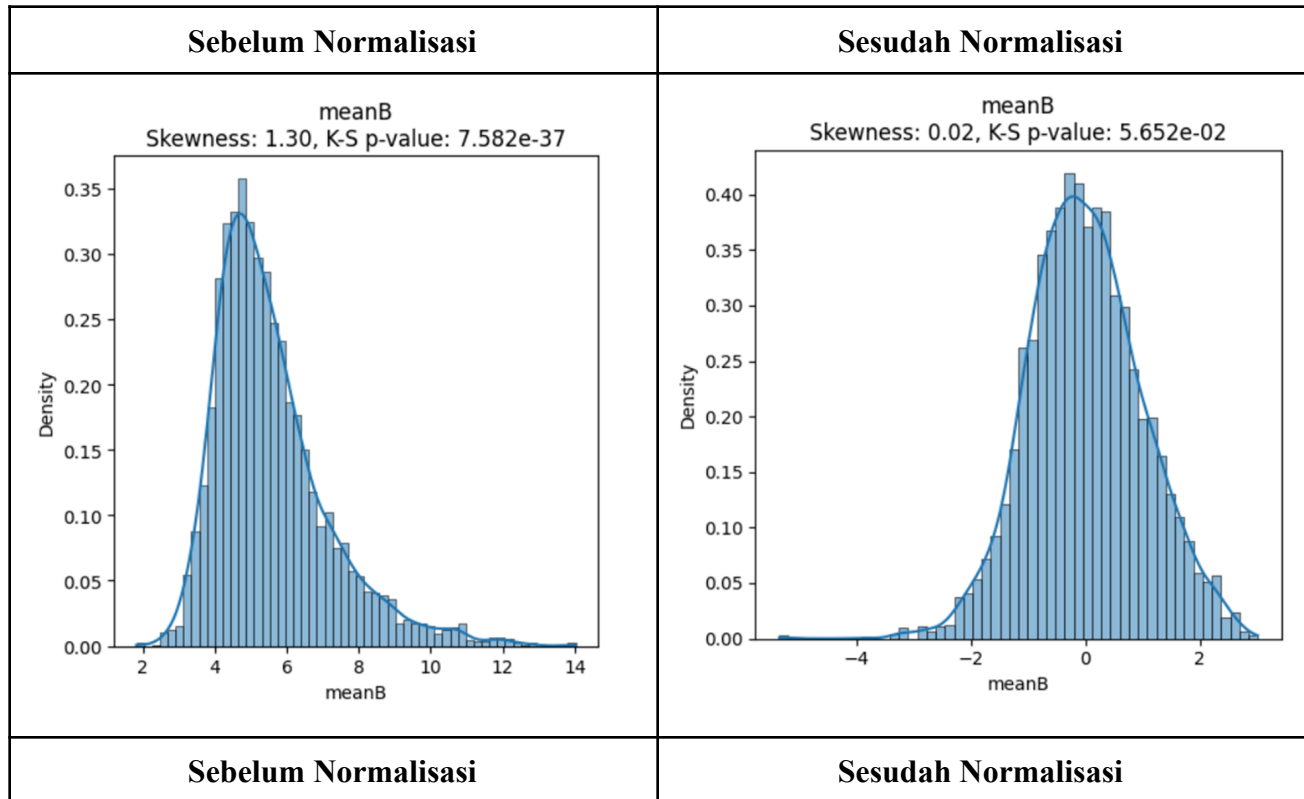
*Yeo-Johnson*, yang merupakan metode yang lebih fleksibel karena dapat menangani nilai negatif dan nol, menunjukkan hasil yang lebih konsisten dalam menormalkan fitur dengan distribusi yang lebih kompleks, seperti meanG dan meanB. *Yeo-Johnson* seringkali lebih unggul karena kemampuan untuk menyesuaikan transformasi pada fitur dengan distribusi non-normal yang lebih beragam. Namun, meskipun metode ini efektif, beberapa fitur masih gagal dinormalisasi dengan baik, terutama yang memiliki distribusi yang sangat miring atau tidak simetris.

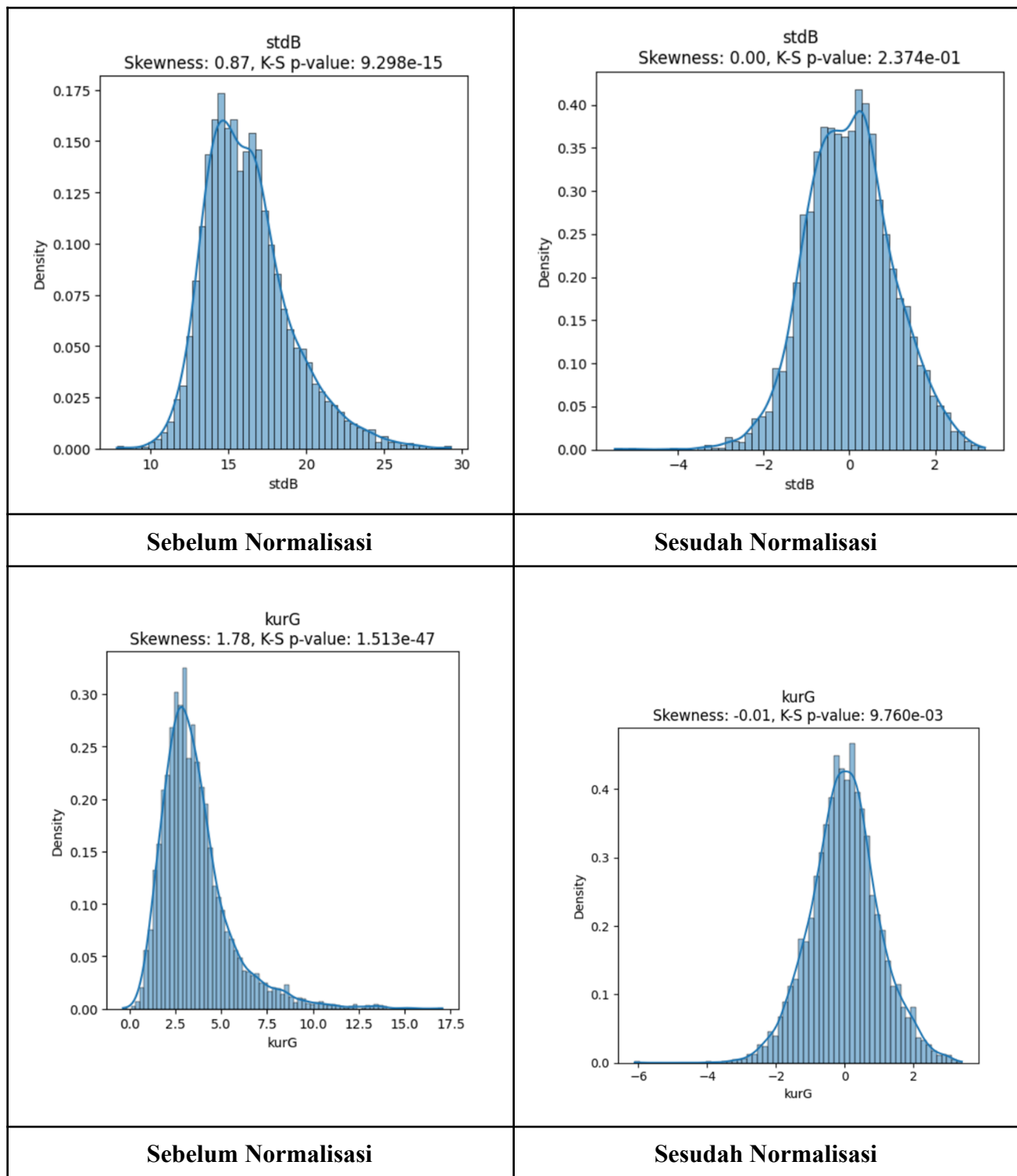
Secara keseluruhan, *Yeo-Johnson* mungkin menjadi metode yang lebih baik karena fleksibilitasnya dalam menangani berbagai jenis distribusi, termasuk yang memiliki nilai negatif atau nol, yang memberikan keuntungan dalam menangani data yang lebih beragam. *Box-Cox* cenderung lebih baik untuk data yang sudah cukup mendekati distribusi normal tetapi tidak dapat menangani distribusi yang sangat miring atau dengan pencilan ekstrem. *Log Test* dan *Square Test* terbatas dalam cakupannya dan sering gagal untuk fitur dengan distribusi yang lebih kompleks, meskipun dapat bekerja dengan baik pada data yang memiliki distribusi lebih sederhana.

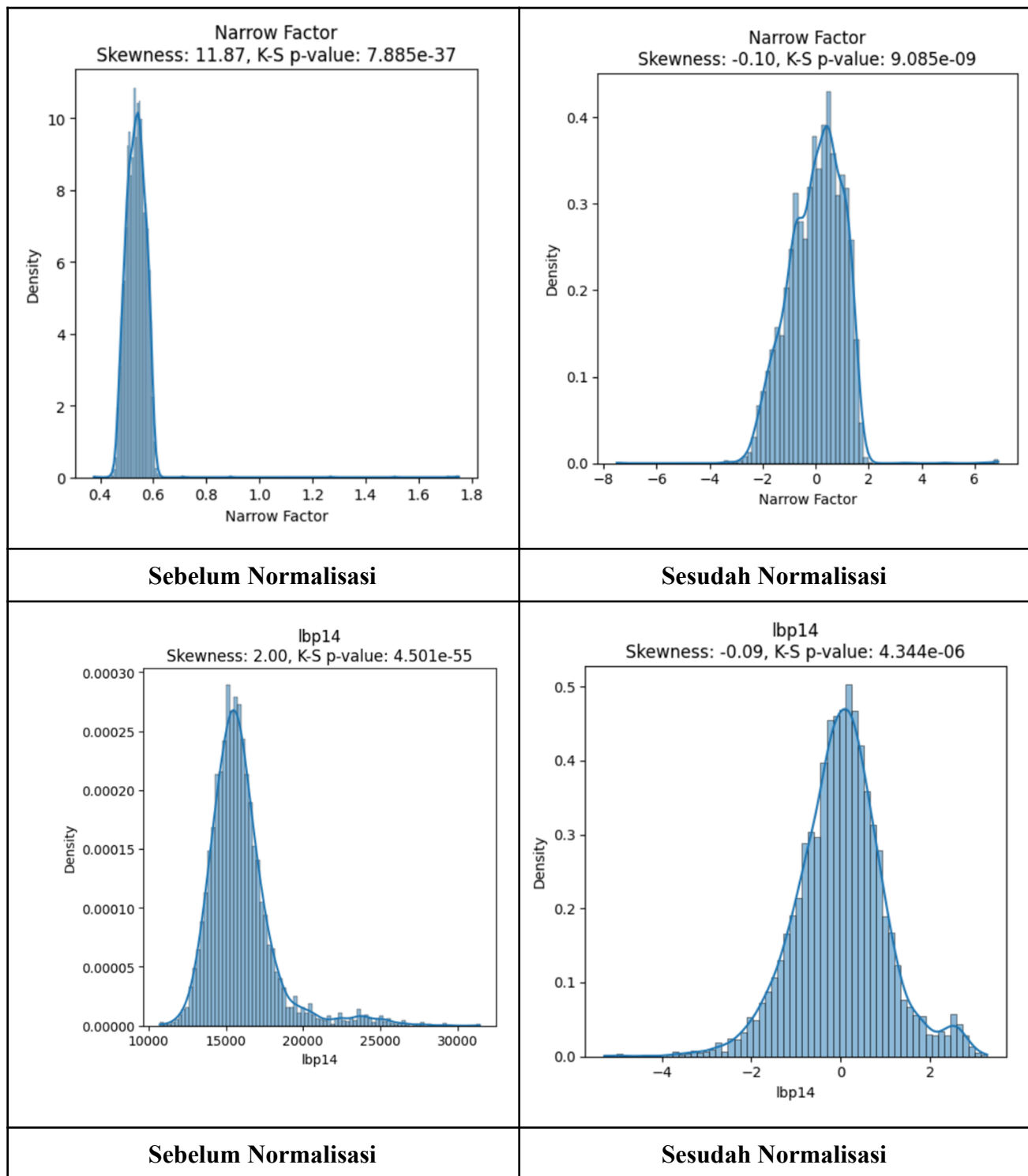
# Visualisasi Setelah Normalisasi

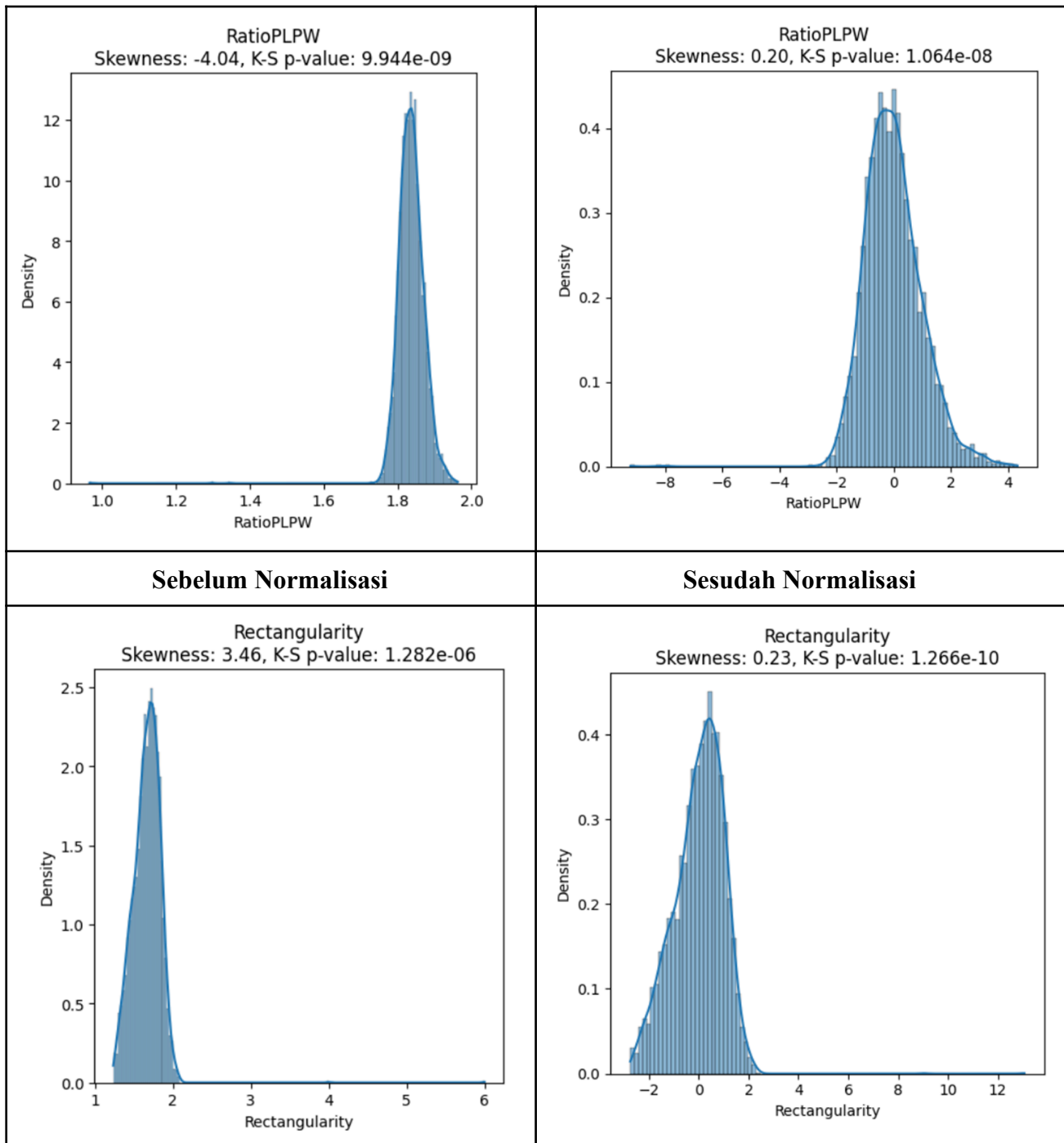
Visualisasi dibawah ini menampilkan data sebelum dan sesudah normalisasi.

**Tabel 4.1 Visualisasi Data Sebelum dan Sesudah di Normalisasi**









Pada tabel perbandingan di atas, terlihat adanya perbedaan yang signifikan antara data sebelum dan sesudah proses normalisasi, di mana normalisasi dilakukan dengan mengombinasikan beberapa metode berbeda untuk mencapai distribusi data yang lebih seragam. Hasil ini ditunjukkan dengan perubahan yang mencolok pada nilai *skewness*, yang mencerminkan tingkat kemiringan distribusi data, serta pada nilai p-value dari uji *Kolmogorov-Smirnov* yang digunakan untuk

mengukur kecocokan data dengan distribusi normal. *Skewness* yang lebih rendah setelah normalisasi menunjukkan bahwa data telah menjadi lebih simetris, sementara *p-value* yang meningkat atau mendekati nilai signifikan pada pengujian *Kolmogorov-Smirnov* menandakan adanya peningkatan kesesuaian data terhadap distribusi normal. Perubahan-perubahan ini mengindikasikan bahwa metode normalisasi yang digunakan berhasil mengurangi outlier dalam data, sehingga analisis statistik selanjutnya dapat dilakukan dengan lebih akurat.