

ANALISIS DAN VISUALISASI DATA PRAKTIKUM (I3)

UJIAN AKHIR SEMESTER

ANALISA DATA BISNIS

STUDI KASUS: EVALUASI TINGKAT STRESS



Disusun Oleh Kelompok 5:

Airlangga Dwi Satrio	082111633003
Nathanael Erlando Putra	082111633025
Shabrina Al Khansa Putri Sulistio	082111633053

Dosen Pengampu

Indah Werdiningsih, S.Si., M.Kom.

198002172005012001

S-1 SISTEM INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS AIRLANGGA

2023

DAFTAR ISI

DAFTAR ISI.....	2
DAFTAR TABEL.....	4
DAFTAR GAMBAR.....	5
PENDAHULUAN.....	6
A. Latar Belakang.....	6
B. Rumusan Masalah.....	7
C. Tujuan Penelitian.....	7
D. Manfaat penelitian.....	8
HASIL DAN PEMBAHASAN.....	9
A. Deskripsi Data.....	9
a. Instrumen Data.....	9
b. Pengumpulan Data.....	10
c. Deskripsi Karakteristik Responden.....	10
d. Deskripsi Jawaban Responden.....	11
B. Data Preprocessing.....	13
a. Missing Value.....	13
b. Outlier.....	13
C. Eksplorasi Data.....	14
a. Boxplot.....	14
b. Histogram.....	15
c. Heat Map.....	16
d. Scatter Plot.....	17
D. Visualisasi Data.....	18
a. Diagram Batang.....	18
b. Line Chart.....	21
c. Pie Chart.....	22
d. Area Chart.....	23
e. Multivariate Visualization.....	25
f. High-Dimensional.....	26
E. Visualizing Classification.....	28
a. Decision Tree.....	29
b. Random Forest.....	30
c. K-Nearest Neighbors.....	32
d. Support Vector Machine.....	32
e. Logistic Regression.....	33
f. Naïve Bayes.....	33
F. Visualisasi kluster analisis (Clustering).....	34

G. Eksplorasi grafis data bisnis GUI.....	35
KESIMPULAN DAN SARAN.....	39
DAFTAR PUSTAKA.....	40

DAFTAR TABEL

Tabel 1. Instrumen Data.....	9
Tabel 2. Frekuensi Berdasarkan Jenis Kelamin Responden.....	10
Tabel 3. Frekuensi Berdasarkan Okupasi Responden.....	10
Tabel 4. Frekuensi Berdasarkan Pendidikan Responden.....	11
Tabel 5. Frekuensi Jawaban Responden.....	11

DAFTAR GAMBAR

Gambar 1. Boxplot Outlier - Jam Tidur Per Hari; Kesibukan Perhari.....	13
Gambar 2. Boxplot Variabel - Jam Tidur Per Hari; Kesibukan Perhari.....	13
Gambar 3. Boxplot Distribusi Umur Berdasarkan Tingkat Stress.....	14
Gambar 4. Histogram frekuensi tingkat stress.....	15
Gambar 5. Heatmap Korelasi.....	16
Gambar 6. Scatter Plot.....	17
Gambar 7. Diagram Batang Distribusi Skor Stress.....	18
Gambar 8. Diagram Batang Tingkat Stress Berdasarkan Jenis Kelamin.....	19
Gambar 9. Diagram Batang Distribusi Frekuensi Status Berdasarkan Tingkat Stress....	20
Gambar 10. Line Chart Tren Tingkat Stress Seiring Waktu.....	21
Gambar 11. Pie Chart Distribusi - Jenis Kelamin; Status; Okupasi; Tingkat Stress.....	22
Gambar 12. Area Chart Tingkat Stress Seiring Waktu.....	23
Gambar 13. Area Chart Kesibukan Per Hari Berdasarkan Score dan Tingkat Stress.....	24
Gambar 14. Pair Plot.....	25
Gambar 15. Bar Plot Kesibukan Per hari Berdasarkan Usia dan Tingkat Stress.....	26
Gambar 16. 2D t-SNE Visualization.....	27
Gambar 17. 3D t-SNE Visualization.....	27
Gambar 18. 3D Pie Chart.....	28
Gambar 19. Perbandingan akurasi model klasifikasi.....	28
Gambar 20. Evaluasi Model Decision Tree.....	29
Gambar 21. Decision Tree Visualization.....	30
Gambar 22. Evaluasi Model Decision Tree.....	30
Gambar 23. Random Forest Visualization.....	31
Gambar 24. Random Forest - Feature Importance.....	32
Gambar 25. Evaluasi Model K-Nearest Neighbors.....	32
Gambar 26. Evaluasi Model Support Vector Machine.....	33
Gambar 27. Evaluasi Model Logistic Regression.....	33
Gambar 28. Evaluasi Model Naive Bayes.....	34
Gambar 29. Elbow Method.....	34
Gambar 30. Anggota Tiap Cluster.....	35
Gambar 31. Persebaran Data.....	35
Gambar 32. GUI Streamlit - Input User.....	37
Gambar 33. GUI Streamlit - Output.....	38

PENDAHULUAN

A. Latar Belakang

Dalam konteks dinamika modern yang diwarnai oleh kemajuan teknologi, persaingan ketat, perubahan gaya hidup yang cepat, dan tuntutan pekerjaan yang semakin kompleks, tingkat stres yang dialami oleh masyarakat umum menjadi sebuah tantangan serius. Stres merupakan kondisi mental yang timbul ketika seseorang menghadapi kesulitan dalam mencapai kebutuhan atau keinginan pribadi, menyebabkan tekanan yang berdampak pada kesehatan mental dan fisik, serta produktivitas kerja individu [3].

Penting untuk memahami dan mengidentifikasi tingkat stress yang dialami oleh masyarakat umum, karena stress yang berlebihan dapat menyebabkan dampak negatif terhadap kesehatan fisik dan mental individu. Respon stres merujuk pada reaksi yang muncul dalam tubuh manusia sebagai tanggapan terhadap tekanan atau tuntutan tertentu yang dikenakan pada tubuh, dan timbul sebagai akibat dari pemicu stres [5]. Beberapa studi menunjukkan bahwa tingkat stres yang tinggi dapat berkontribusi pada peningkatan risiko penyakit jantung, gangguan tidur, gangguan pencernaan, dan bahkan penurunan sistem kekebalan tubuh. Oleh karena itu, pemahaman mendalam tentang tingkat stres masyarakat bukan hanya relevan untuk kesehatan individu, tetapi juga untuk kesejahteraan masyarakat secara keseluruhan.

Dalam penelitian ini, pengukuran tingkat stres masyarakat akan menggunakan pendekatan yang dapat diukur secara empiris, salah satunya dengan memanfaatkan *Perceived Stress Scale* (PSS). PSS adalah alat pengukuran yang telah terbukti valid dan reliabel dalam mengukur tingkat stres yang dirasakan oleh individu. PSS menilai sejauh mana responden menganggap hidup ini tidak dapat diprediksi, tidak dapat dikendalikan, dan membebani [1]. Alat ini mengajukan pertanyaan terstruktur untuk menilai sejauh mana seseorang merasakan situasi hidupnya sebagai pemicu stres, dengan rentang waktu tertentu. Penggunaan PSS dalam penelitian ini akan memberikan dimensi kuantitatif yang lebih jelas terkait tingkat stres masyarakat.

Dalam penelitian ini, metode clustering diadopsi untuk mengidentifikasi pola-pola tingkat stres yang mungkin tidak terlihat secara langsung, memungkinkan pengelompokan masyarakat ke dalam kategori-kategori berdasarkan karakteristik stres.

Pendekatan ini memberikan dasar yang kuat untuk pengembangan solusi yang lebih tepat dan terarah. Clustering, sebagai alat yang berguna dalam ilmu data, digunakan untuk menemukan struktur cluster dalam kumpulan data, yang ditandai dengan kesamaan terbesar dalam cluster yang sama dan ketidaksamaan terbesar antara cluster yang berbeda [4].

Hasil dari penelitian ini diharapkan dapat memberikan landasan yang kuat untuk perumusan kebijakan kesehatan mental dan pengembangan intervensi yang lebih efektif, dengan mempertimbangkan data PSS sebagai salah satu indikator tingkat stres. Dengan pemahaman yang lebih baik tentang faktor-faktor pemicu stres, kebijakan dapat dirancang untuk mendukung kesejahteraan masyarakat secara menyeluruh, mengurangi dampak negatif stres terhadap individu, dan memastikan solusi yang sesuai dengan konteks budaya dan sosial masyarakat. Melalui pendekatan komprehensif ini, penelitian ini diharapkan memberikan kontribusi yang signifikan untuk pemahaman tingkat stres masyarakat dan menyediakan dasar yang kokoh untuk solusi yang relevan dan efektif.

B. Rumusan Masalah

Penelitian ini bertujuan untuk menjawab pertanyaan-pertanyaan utama berikut:

1. Bagaimana tingkat stres yang dialami oleh masyarakat dalam konteks dinamika modern?
2. Apa saja faktor-faktor pemicu stres yang dominan dalam masyarakat?
3. Bagaimana pola-pola tingkat stres dapat diidentifikasi melalui pendekatan clustering?

C. Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengukur tingkat stres masyarakat menggunakan Perceived Stress Scale (PSS) sebagai alat pengukuran empiris.
2. Mengidentifikasi faktor-faktor pemicu stres yang signifikan dalam konteks modern.
3. Menerapkan metode clustering untuk mengelompokkan masyarakat berdasarkan karakteristik stres.

D. Manfaat penelitian

Hasil penelitian ini diharapkan memberikan manfaat signifikan, antara lain dengan menyediakan landasan bagi perumusan kebijakan kesehatan mental yang lebih efektif, mengembangkan intervensi yang sesuai dengan tingkat stres masyarakat, memberikan kontribusi pada pemahaman faktor-faktor pemicu stres untuk merancang kebijakan yang mendukung kesejahteraan masyarakat, serta menyediakan dasar bagi solusi yang relevan dan efektif dalam konteks budaya dan sosial masyarakat. Dengan demikian, penelitian ini diharapkan dapat secara holistik meningkatkan pemahaman terhadap tingkat stres masyarakat dan mendukung upaya-upaya yang bertujuan pada kesejahteraan mental dan fisik individu serta masyarakat secara keseluruhan.

HASIL DAN PEMBAHASAN

A. Deskripsi Data

Deskripsi data adalah gambaran data yang digunakan dalam penelitian. Dari pengumpulan data kuesioner, jumlah responden yang didapatkan sebanyak 156 orang dengan persebaran yang bervariasi. Dalam deskripsi data ini peneliti akan menjabarkan gambaran atau kondisi responden yang menjadi sampel dalam penelitian ini.

a. Instrumen Data

Instrumen data yang digunakan dalam pengumpulan data dapat dilihat dalam **Tabel 1** berikut.

Tabel 1. Instrumen Data

Aspek	Instrumen
<i>Frequency of Anger</i>	Dalam sebulan terakhir, seberapa sering Anda kesal karena suatu hal terjadi secara tak terduga?
<i>Lack of Control</i>	Dalam sebulan terakhir, seberapa sering Anda merasa tidak mampu mengendalikan hal-hal penting dalam hidup Anda?
<i>Anxiety and Stress</i>	Dalam sebulan terakhir, seberapa sering Anda merasa gugup dan stres?
<i>Self-Efficacy</i>	Dalam sebulan terakhir, seberapa sering Anda merasa yakin dengan kemampuan Anda dalam menangani masalah pribadi?
<i>Perceived Control</i>	Dalam sebulan terakhir, seberapa sering Anda merasa segalanya berjalan sesuai keinginan Anda?
<i>Perceived Overwhelm</i>	Dalam sebulan terakhir, seberapa sering Anda menyadari bahwa Anda tidak mampu menyelesaikan semua hal yang harus Anda lakukan?
<i>Feeling of Mastery</i>	Dalam sebulan terakhir, seberapa sering Anda merasa berada di puncak segalanya?

<i>Cumulative Difficulty</i>	Dalam sebulan terakhir, seberapa sering Anda merasakan kesulitan yang menumpuk sehingga Anda tidak mampu mengatasinya?
------------------------------	--

b. Pengumpulan Data

Pengumpulan data dilakukan dengan menyebarkan kuesioner kepada khalayak luas melalui media sosial menggunakan media *Google Form*.

c. Deskripsi Karakteristik Responden

Berdasarkan data hasil kuesioner diketahui gambaran karakteristik responden yang menjadi sampel dalam penelitian ini, yaitu: berdasarkan jenis kelamin, berdasarkan okupasi, dan berdasarkan pendidikan responden. Hasil lebih rinci disajikan dalam **Tabel 2**.

Tabel 2. Frekuensi Berdasarkan Jenis Kelamin Responden

No.	Jenis Kelamin	Frekuensi (orang)	Persentase
1	Laki-laki	55	35,3%
2	Perempuan	101	64,7%
Jumlah		156	100%

Berdasarkan **Tabel 2**, terlihat bahwa karakteristik responden berdasarkan jenis kelamin didominasi oleh perempuan sebanyak 101 orang dari keseluruhan 156 responden.

Tabel 3. Frekuensi Berdasarkan Okupasi Responden

No.	Okupasi	Frekuensi (orang)	Persentase
1	Bekerja	22	14,1%
2	Tidak Bekerja	134	85,9%
Jumlah		156	100%

Berdasarkan **Tabel 3**, terlihat bahwa karakteristik responden berdasarkan okupasi sebanyak 85,9% dari keseluruhan 156 responden tidak bekerja.

Tabel 4. Frekuensi Berdasarkan Pendidikan Responden

No.	Pendidikan	Frekuensi (orang)	Persentase
1	Pelajar	5	3,2%
2	Mahasiswa	144	92,3%
3	Lainnya	7	4,5%
Jumlah		156	100%

Berdasarkan **Tabel 4**, terlihat bahwa karakteristik responden berdasarkan pendidikan didominasi oleh mahasiswa sebanyak 144 orang dari keseluruhan 156 responden.

d. Deskripsi Jawaban Responden

Hasil data dari kuesioner yang sudah disebar menggunakan media *Google Form* dapat dilihat pada **Tabel 5**.

Tabel 5. Frekuensi Jawaban Responden

Instrumen	Jawaban (skor)									
	1		2		3		4		5	
	F	%	F	%	F	%	F	%	F	%
Dalam sebulan terakhir, seberapa sering Anda kesal karena suatu hal terjadi secara tak terduga?	2	1,28	10	6,41	53	33,97	54	34,61	37	23,71
Dalam sebulan terakhir, seberapa sering Anda merasa tidak mampu mengendalikan hal-hal penting	1	0,64	21	13,46	48	30,76	53	33,97	33	21,15

Instrumen	Jawaban (skor)									
	1		2		3		4		5	
	F	%	F	%	F	%	F	%	F	%
dalam hidup Anda?										
Dalam sebulan terakhir, seberapa sering Anda merasa gugup dan stres?	2	1,28	13	8,33	33	21,15	67	42,94	40	26,28
Dalam sebulan terakhir, seberapa sering Anda merasa yakin dengan kemampuan Anda dalam menangani masalah pribadi?	1	0,64	38	24,35	60	38,46	40	25,64	17	10,89
Dalam sebulan terakhir, seberapa sering Anda merasa segalanya berjalan sesuai keinginan Anda?	13	8,33	47	30,12	61	39,10	31	19,87	4	2,56
Dalam sebulan terakhir, seberapa sering Anda menyadari bahwa Anda tidak mampu menyelesaikan semua hal yang harus Anda lakukan?	1	0,64	28	17,94	48	30,76	59	37,82	20	12,83
Dalam sebulan terakhir, seberapa sering Anda merasa berada di puncak segalanya?	29	18,58	49	31,41	51	32,69	18	11,53	9	5,76
Dalam sebulan terakhir, seberapa sering Anda merasakan kesulitan yang menumpuk sehingga Anda tidak mampu mengatasinya?	3	1,92	26	16,66	35	22,43	48	30,76	44	28,20

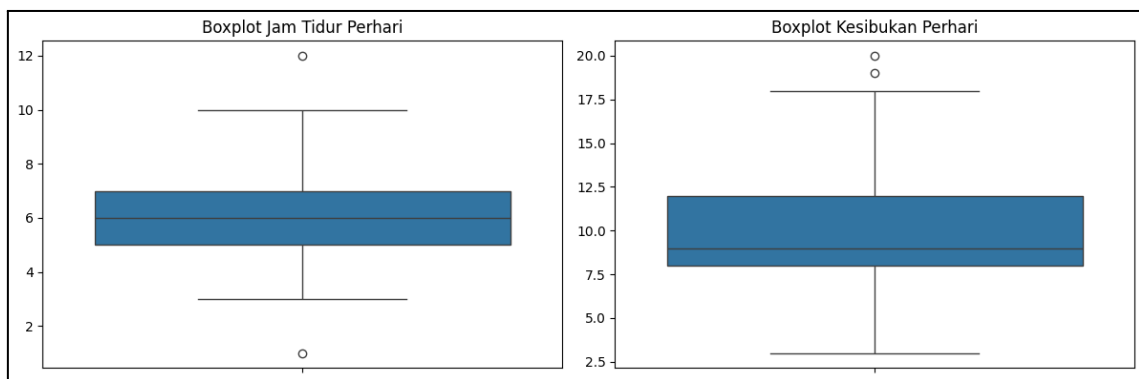
B. Data Preprocessing

a. Missing Value

Dalam sampel data yang digunakan tidak terdapat *missing value* sehingga tidak memerlukan proses penanganan untuk *missing value*.

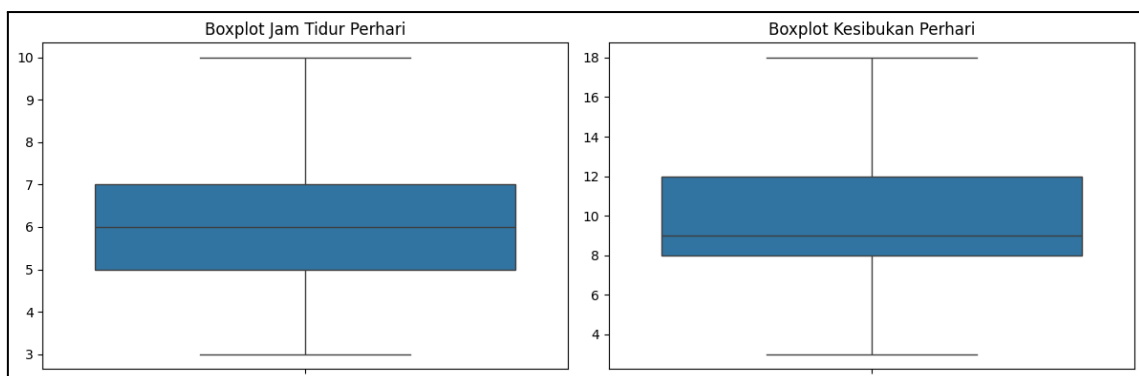
b. Outlier

Berdasarkan sampel data yang digunakan, terdapat *outlier* pada variabel jam tidur perhari dan variabel kesibukan perhari yang ditunjukkan menggunakan boxplot. Visualisasi *outlier* dapat dilihat pada **Gambar 1**.



Gambar 1. Boxplot Outlier - Jam Tidur Per Hari; Kesibukan Perhari

Dalam **Gambar 1** terlihat bahwa terdapat nilai 12 pada variabel jam tidur perhari yang merupakan *outlier*; sedangkan pada variabel kesibukan perhari terdapat nilai 18 dan 20 yang merupakan *outlier*.

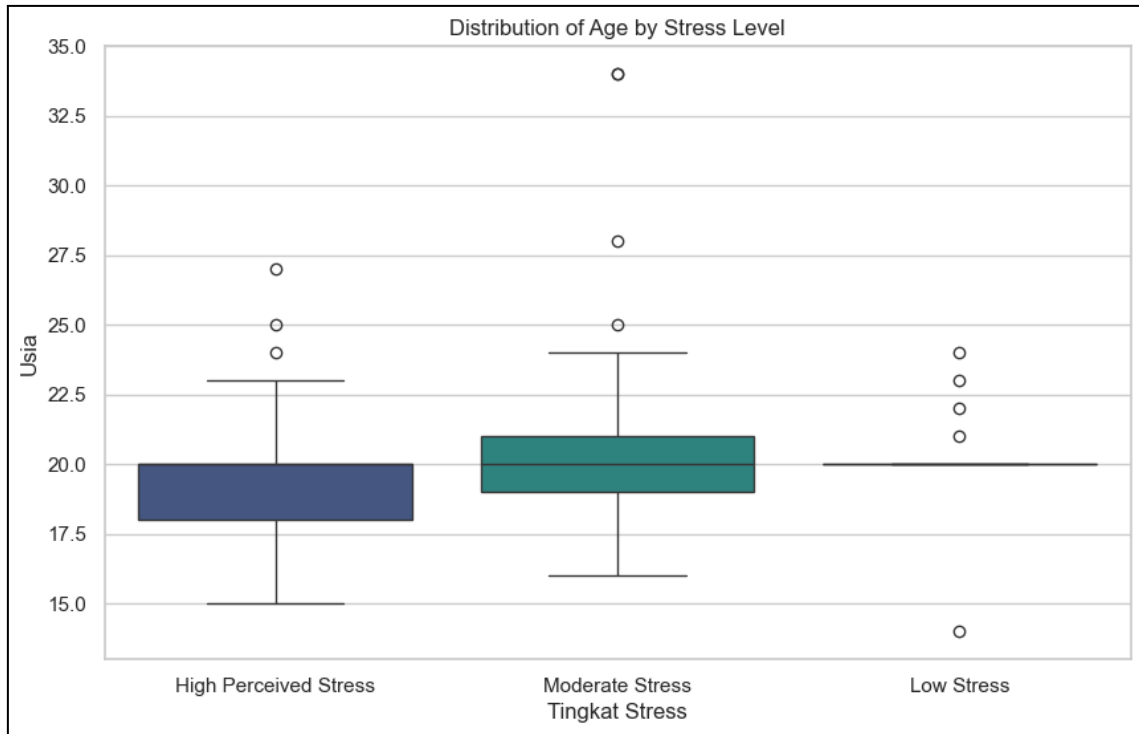


Gambar 2. Boxplot Variabel - Jam Tidur Per Hari; Kesibukan Perhari

Penanganan yang dilakukan untuk mengatasi *outlier* adalah menggantinya dengan nilai *mean*. Hasil penanganan *outlier* dapat dilihat dalam **Gambar 2**.

C. Eksplorasi Data

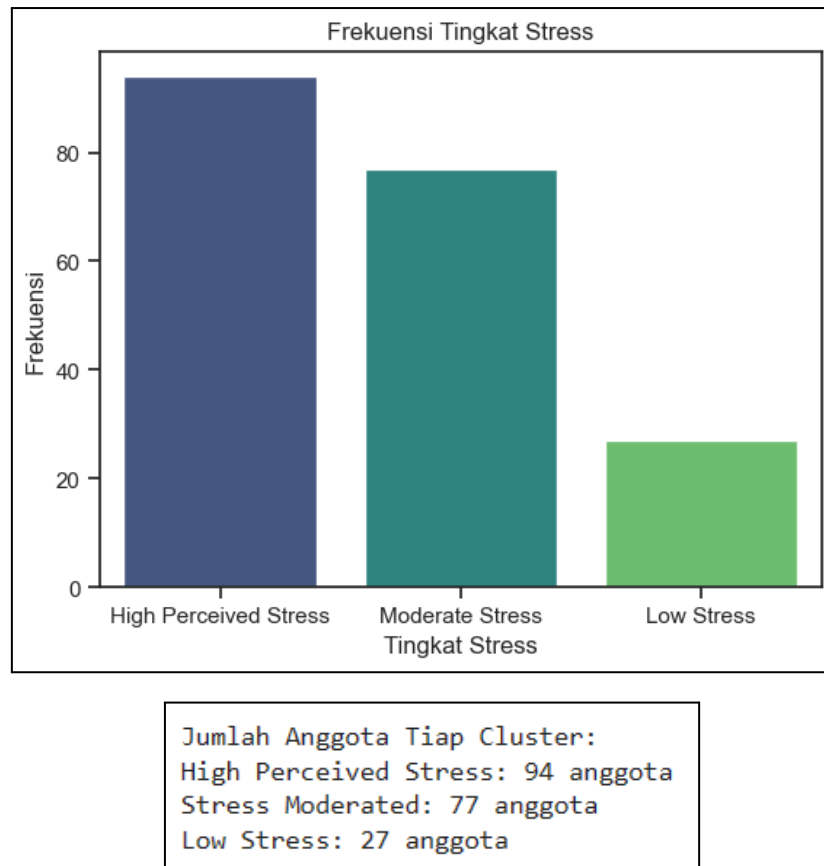
a. Boxplot



Gambar 3. Boxplot Distribusi Umur Berdasarkan Tingkat Stress

Berdasarkan hasil visualisasi boxplot pada **Gambar 3**, terlihat bahwa data observasi terkecil tingkat stress *high perceived* adalah 15 tahun dengan data observasi tertinggi 26 tahun, data observasi terkecil tingkat stress *moderated* adalah 16 tahun dengan data observasi tertinggi 34 tahun, dan data observasi terkecil tingkat stress *low* adalah 13 tahun dengan data observasi tertinggi 24 tahun.

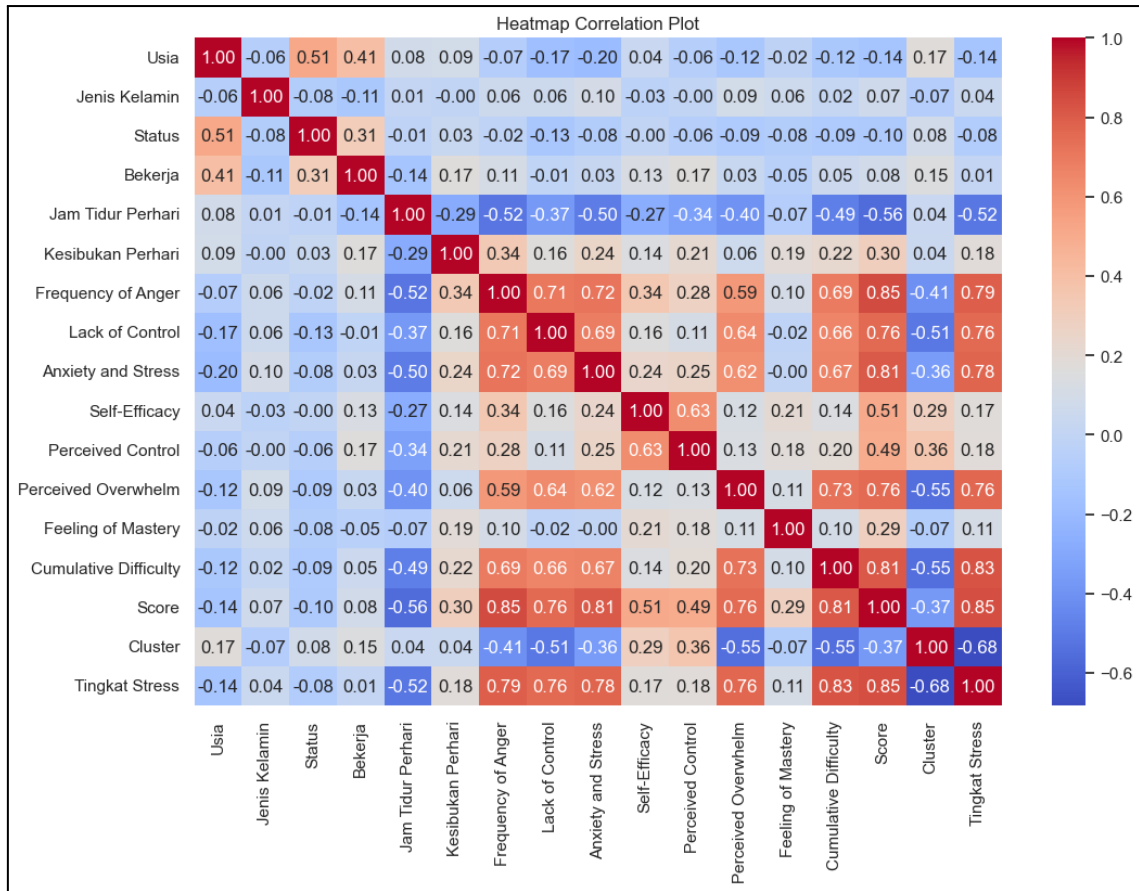
b. Histogram



Gambar 4. Histogram frekuensi tingkat stress

Distribusi frekuensi tingkat stress pada sampel data yang ditunjukkan oleh **Gambar 4** menunjukkan frekuensi tingkat stress tertinggi adalah *high perceived stress* dengan jumlah anggota sebanyak 94, diikuti dengan *moderate stress* dengan jumlah anggota sebanyak 77, dan frekuensi terendah adalah *low stress* dengan jumlah anggota sebanyak 27.

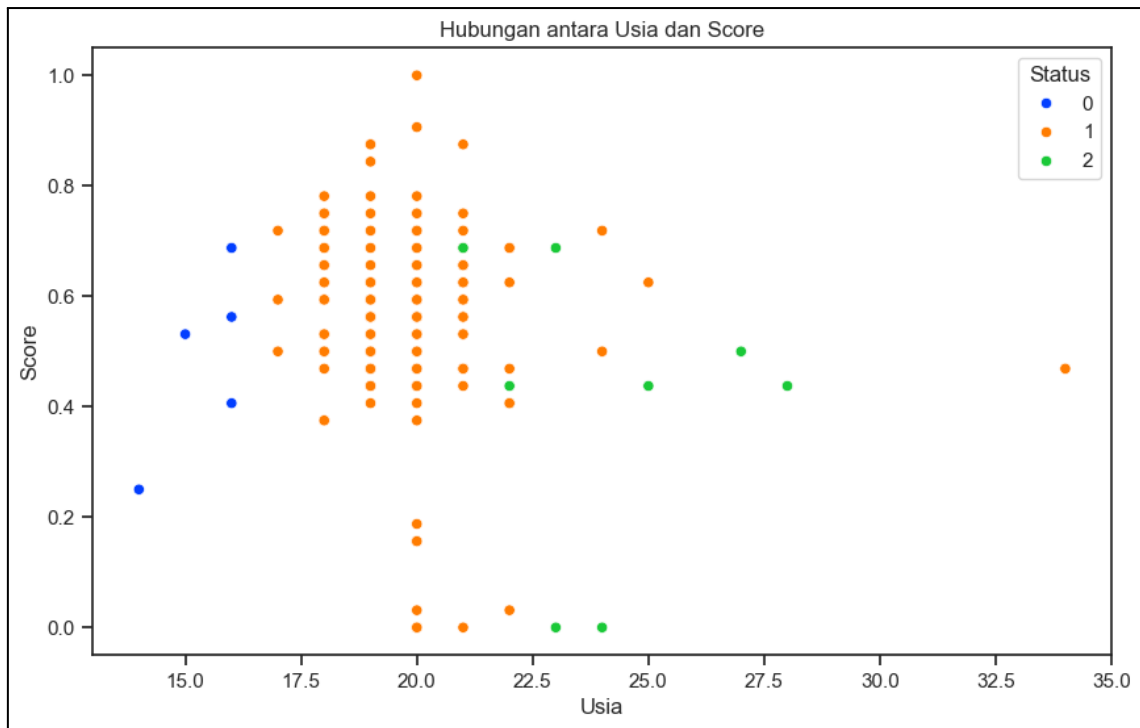
c. Heat Map



Gambar 5. Heatmap Korelasi

Heatmap korelasi pada **Gambar 5** menunjukkan bahwa korelasi tertinggi adalah antara variabel Skor dengan variabel tingkat stress dengan nilai 0.85, disusul dengan korelasi variabel *cumulative difficulty* dengan variabel tingkat stress yang bernilai 0.83, dan korelasi variabel tingkat stress dengan variabel *frequency of anger* yang bernilai 0.79.

d. Scatter Plot



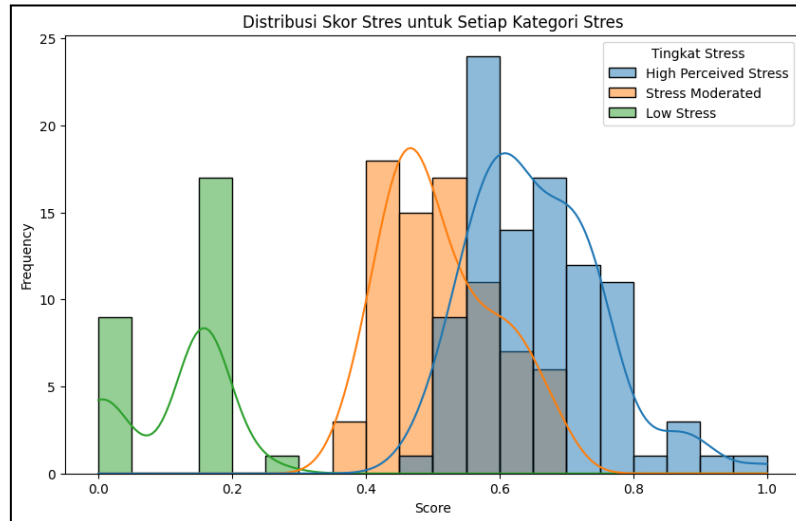
Gambar 6. Scatter Plot

Hasil visualisasi **Gambar 6** menunjukkan hubungan antara usia, skor, dan status. Mayoritas data berusia 17-25 dengan skor 0.4-1.0 berstatus 1 atau mahasiswa, sedangkan data berusia di atas 22.5 dengan skor yang bervariasi berstatus 2 atau lainnya.

D. Visualisasi Data

a. Diagram Batang

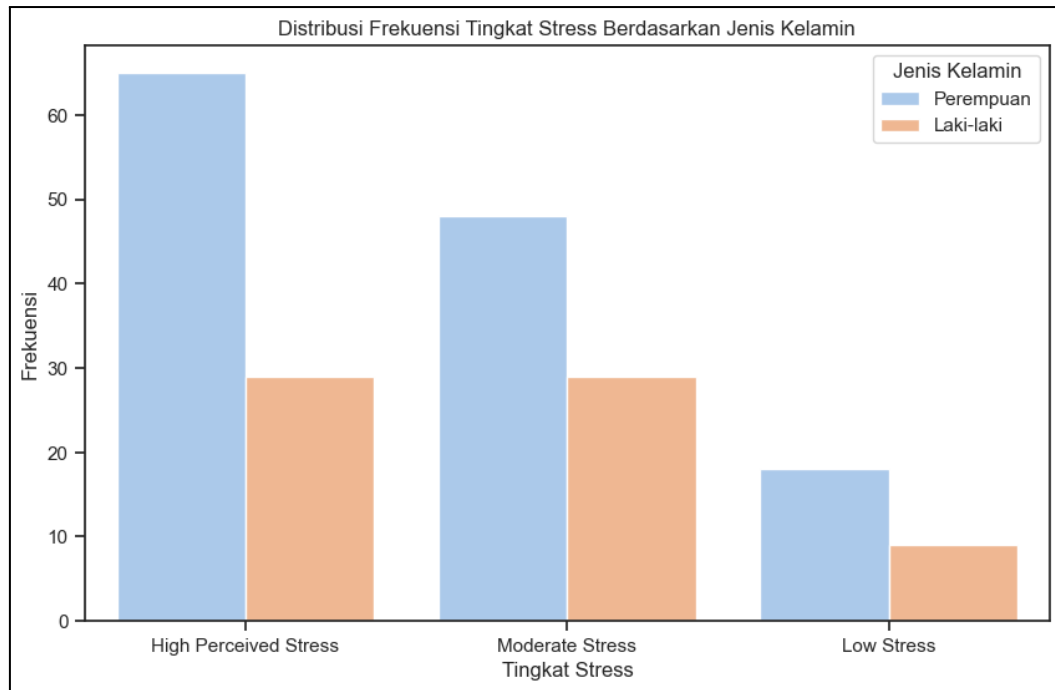
i. Distribusi Skor Stress untuk Setiap Kategori Stress



Gambar 7. Diagram Batang Distribusi Skor Stress

Berdasarkan hasil visualisasi **Gambar 7**, frekuensi tertinggi dalam sampel data adalah *high perceived stress* dengan skor sebesar 0.55-0.6, disusul dengan *stress moderated* dengan skor 0.4-0.45. Distribusi terbanyak pada sampel data adalah *high perceived stress*.

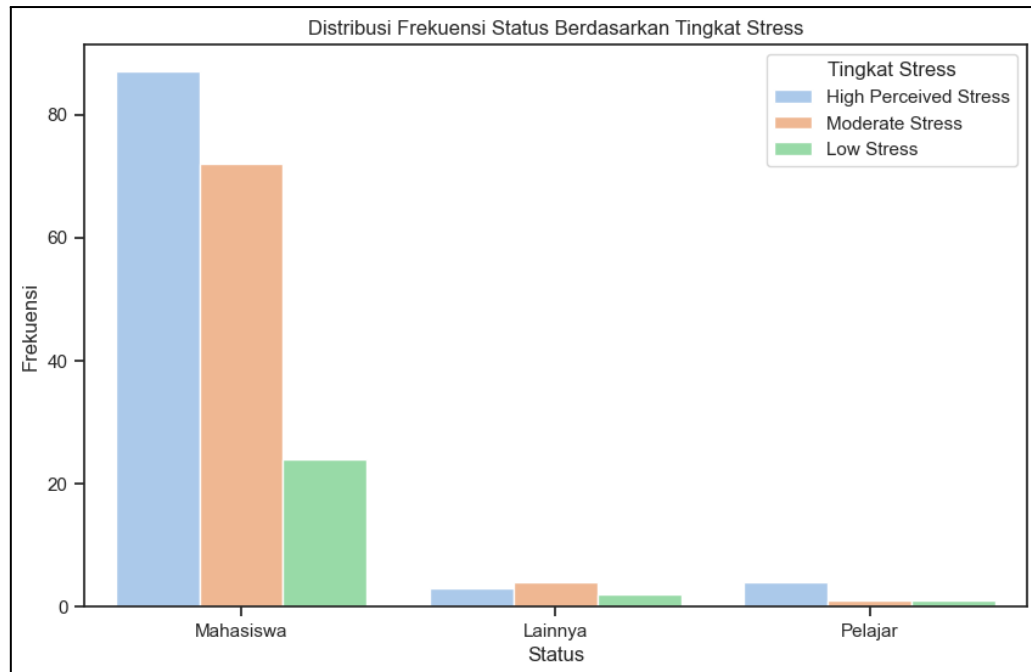
ii. Distribusi Frekuensi Tingkat Stress berdasarkan Jenis Kelamin



Gambar 8. Diagram Batang Tingkat Stress Berdasarkan Jenis Kelamin

Distribusi frekuensi tingkat stress berdasarkan jenis kelamin pada tingkat *high perceived stress*, *moderate stresss*, dan *low stress*, ketiganya didominasi oleh perempuan. Berdasarkan **Gambar 8** terlihat bahwa frekuensi laki-laki dengan tingkat stress *high perceived* dan *moderated* sama yaitu sebanyak 30 orang masing-masing.

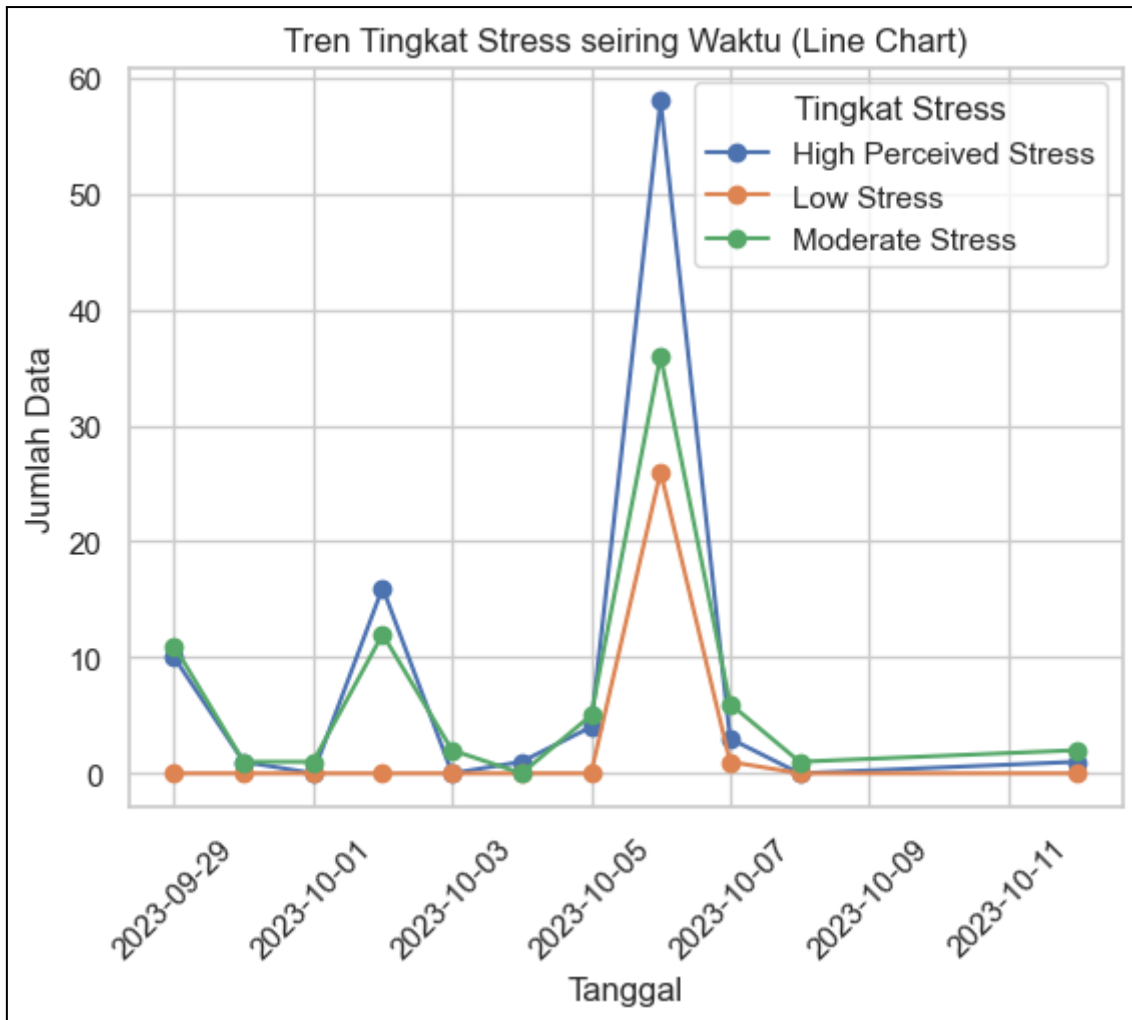
iii. Distribusi Frekuensi Status berdasarkan Tingkat Stress



Gambar 9. Diagram Batang Distribusi Frekuensi Status Berdasarkan Tingkat Stress

Berdasarkan hasil visualisasi **Gambar 9**, frekuensi status berdasarkan tingkat stress didominasi oleh mahasiswa dengan *high perceived stress* sebanyak 90, *moderate stress* sebanyak 70, dan *low stress* sebanyak 20, sedangkan frekuensi pelajar dan lainnya tidak mencapai 20.

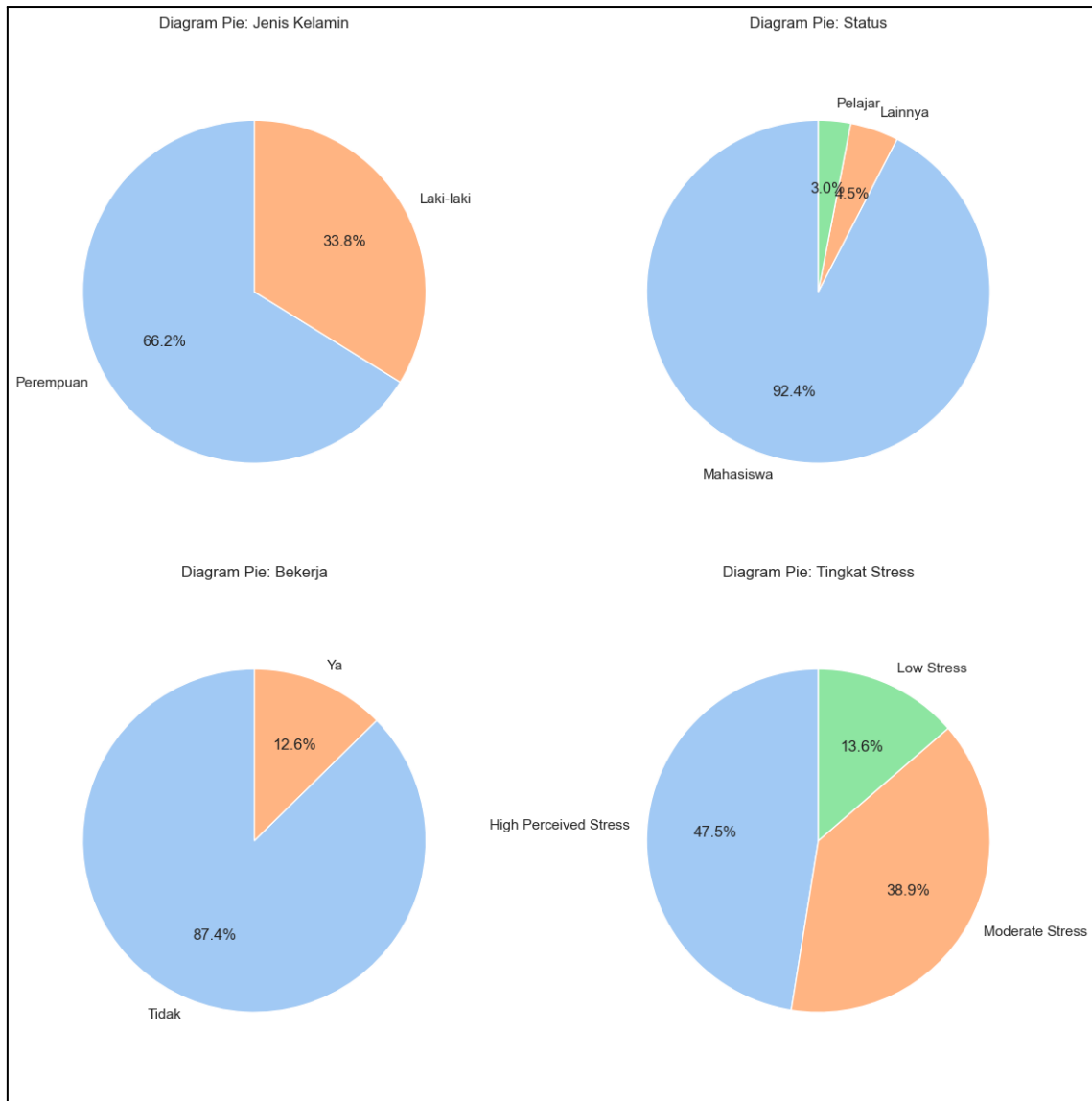
b. Line Chart



Gambar 10. Line Chart Tren Tingkat Stress Seiring Waktu

Tren tingkat stress seiring waktu yang disajikan dalam **Gambar 10**, menunjukkan bahwa jumlah data terbanyak didapatkan pada rentang tanggal 5-7 Oktober 2023 dengan jumlah data *high perceived stress* mencapai 60 data, *moderate stress* mencapai 35 data, dan *low stress* mencapai 25 data.

c. Pie Chart

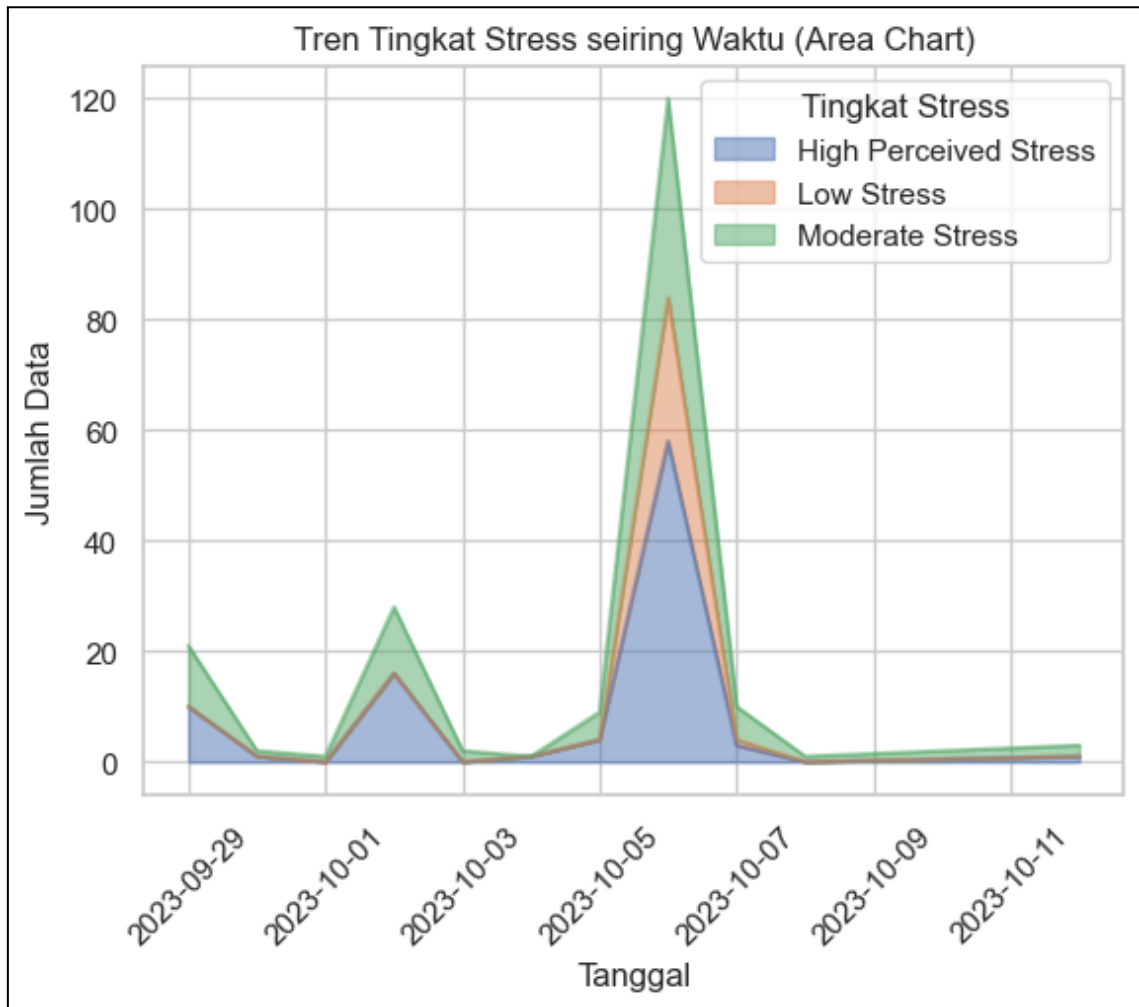


Gambar 11. Pie Chart Distribusi - Jenis Kelamin; Status; Okupasi; Tingkat Stress

Distribusi sampel data berdasarkan jenis kelamin didominasi oleh perempuan dengan persentase sebesar 66,2%. Mayoritas sampel data berdasarkan status pendidikan dan okupasi berturut-turut adalah mahasiswa dan tidak bekerja, dengan nilai 92,4% dan 87,4%, sedangkan distribusi tertinggi sampel data berdasarkan tingkat stress adalah *high perceived stress* dengan persentase sebesar 47,5%. Visualisasi distribusi dapat dilihat pada **Gambar 11**.

d. Area Chart

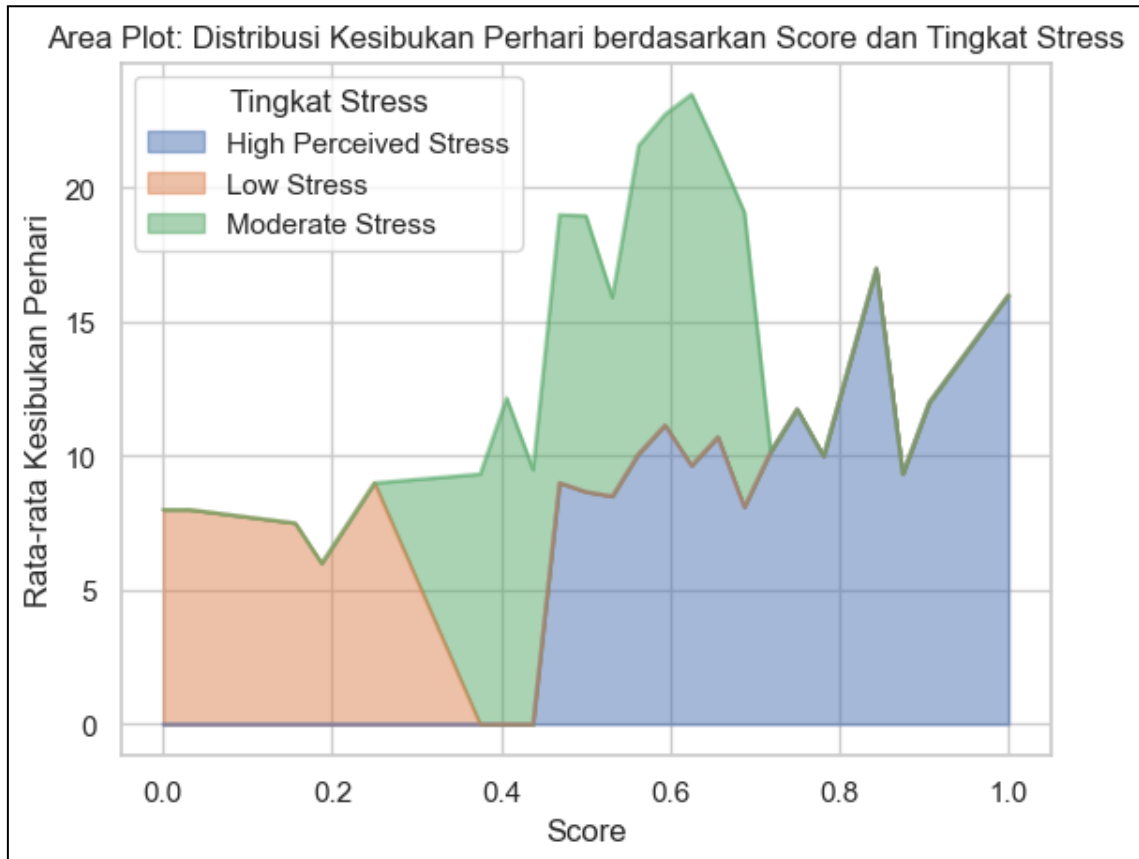
i. Tren Tingkat Stress Seiring Waktu



Gambar 12. Area Chart Tingkat Stress Seiring Waktu

Tren data yang ditunjukkan oleh **Gambar 12** menampilkan jumlah data terbanyak didapat pada rentang tanggal 5-7 Oktober 2023, disusul pada rentang tanggal 1-3 Oktober 2023, sedangkan tren terendah didapat pada rentang tanggal 3-5 Oktober 2023.

ii. Distribusi Kesibukan Perhari Berdasarkan Score dan Tingkat Stress

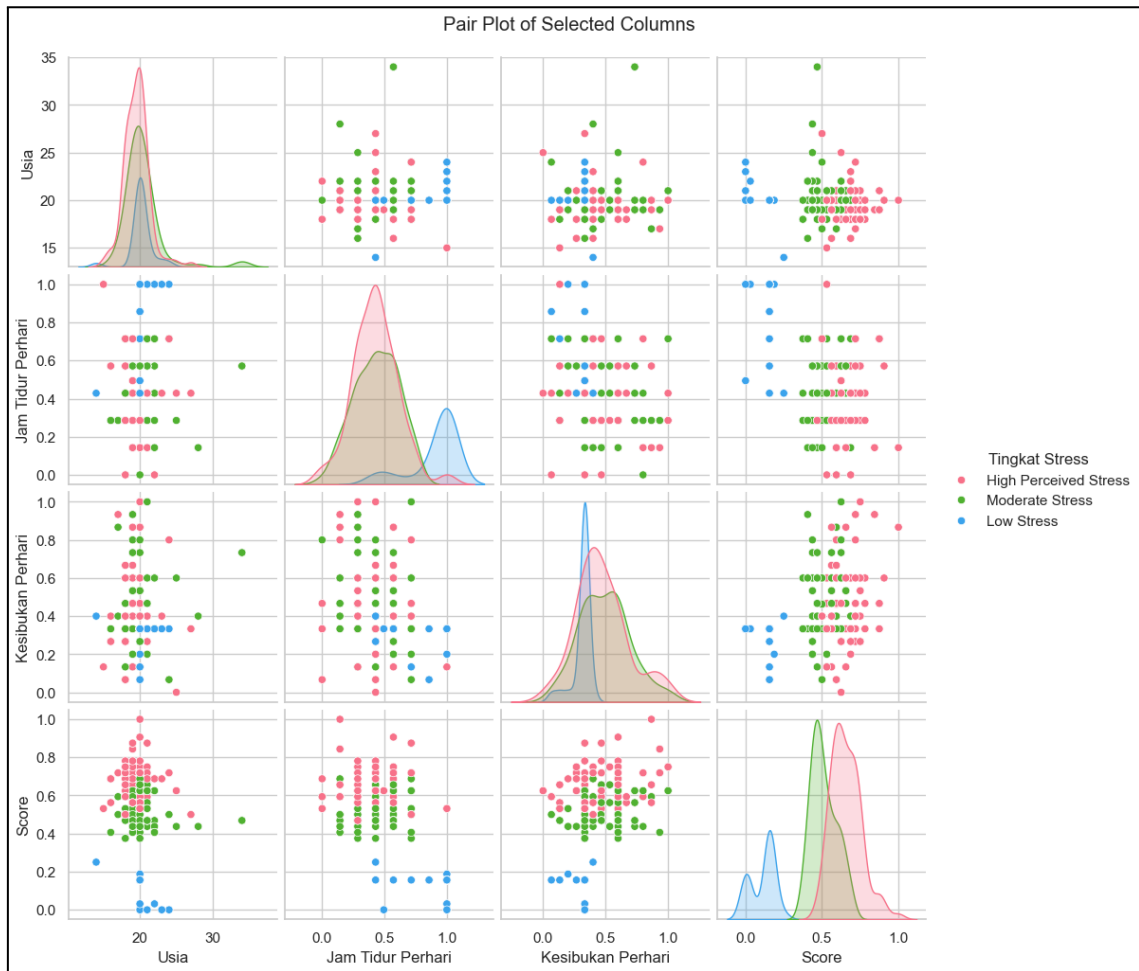


Gambar 13. Area Chart Kesibukan Per Hari Berdasarkan Score dan Tingkat Stress

Berdasarkan **Gambar 13**, rata-rata kesibukan perhari dan skor tidak secara langsung berpengaruh secara linier tingkat stress. Hal ini dinyatakan dengan sampel dengan skor 0.6 dan rata-rata kesibukan perhari yang bervariasi dapat memiliki 2 tingkat stress yaitu *high perceived* dan *moderate*. **Gambar 13** juga menunjukkan bahwa semakin tinggi rata-rata kesibukan perhari belum tentu memiliki tingkat stress yang semakin tinggi pula.

e. Multivariate Visualization

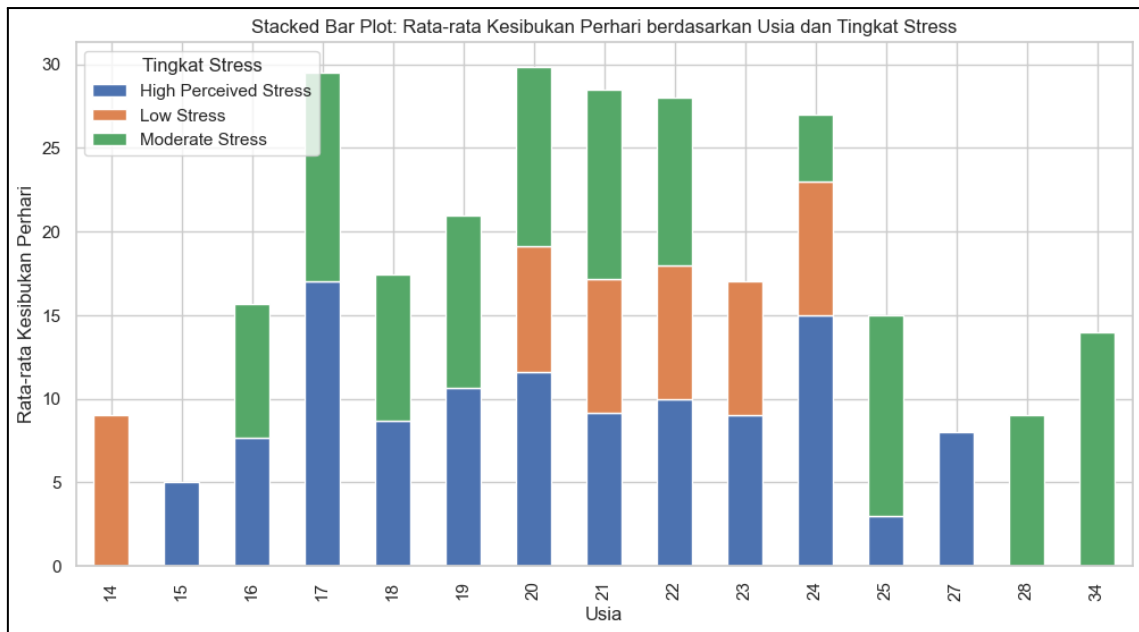
i. Pair Plot



Gambar 14. Pair Plot

Pair plot yang disajikan dalam **Gambar 14** menunjukkan beberapa hubungan antara beberapa variabel. Hubungan antara variabel skor dengan usia menunjukkan bahwa data berusia 20 tahun memiliki nilai skor yang bervariasi, dalam rentang 0.4-1, dengan semakin tinggi nilai skor, semakin tinggi pula tingkat stress. Lalu untuk variabel jam tidur perhari dengan skor juga memiliki hubungan semakin tinggi nilai skor, semakin tinggi pula tingkat stress.

ii. Bar Plot

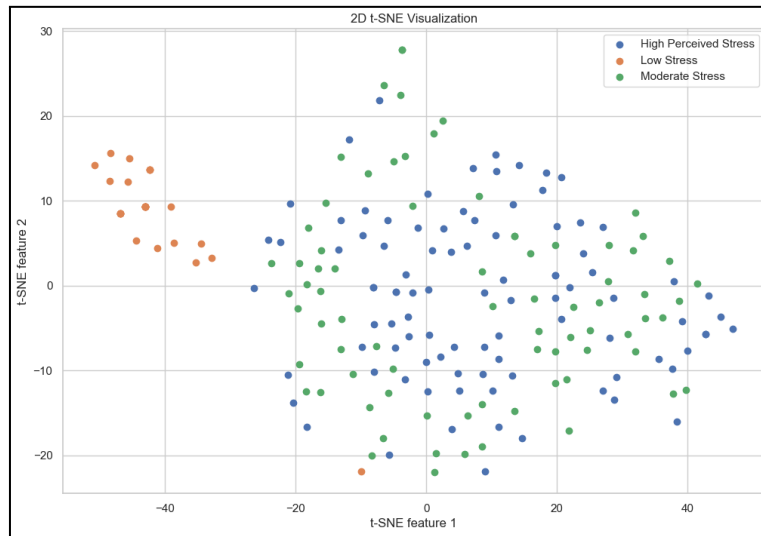


Gambar 15. Bar Plot Kesibukan Per hari Berdasarkan Usia dan Tingkat Stress

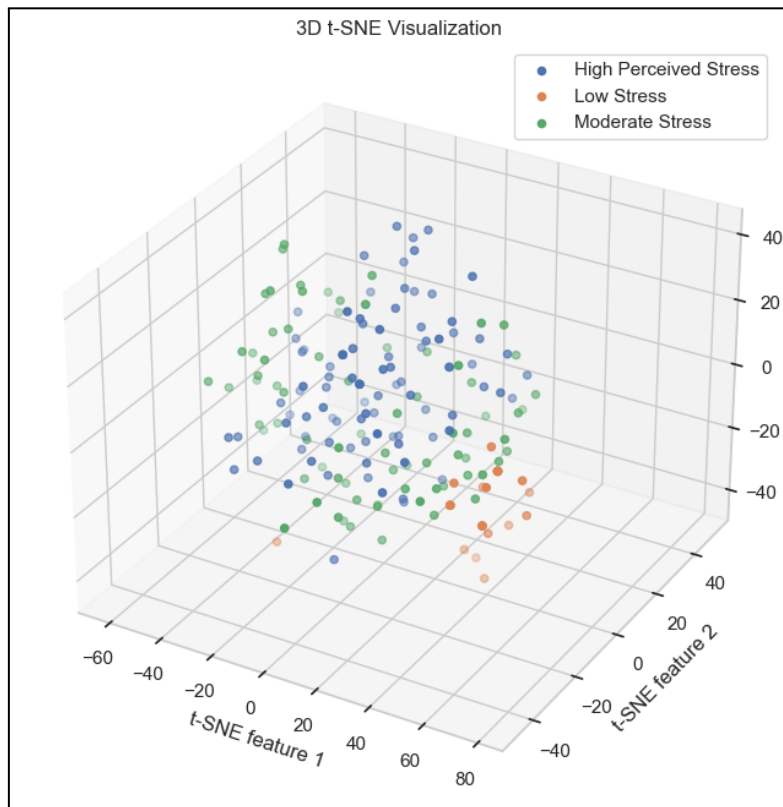
Berdasarkan hasil visualisasi **Gambar 15**, data dengan rata-rata kesibukan perhari tertinggi memiliki tingkat stress *moderate* terlepas dari usia. Hal ini ditunjukkan pada data dengan usia 17 dan 20 yang memiliki rata-rata kesibukan perhari senilai 28 termasuk dalam tingkat stress yang sama yaitu *moderate*.

f. High-Dimensional

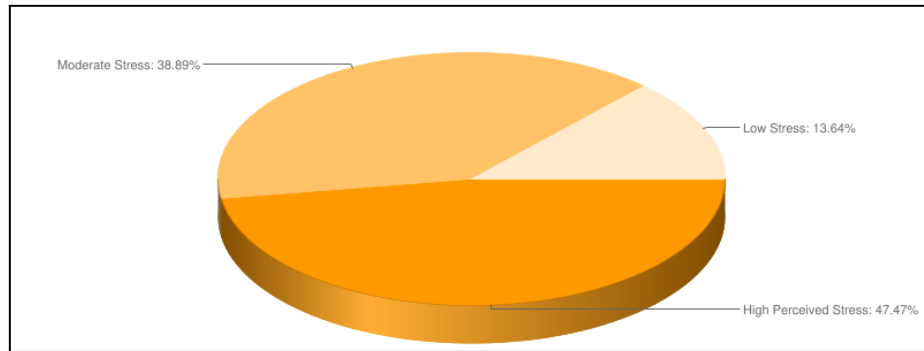
Menggunakan t-SNE(t-Distributed Stochastic Neighbor Embedding) untuk mengurangi data yang berdimensi tinggi. T-Distributed Stochastic Neighbor Embedding memiliki nilai penerapan yang besar dalam *big data*, *data mining*, *machine learning*, *deep learning*, dan bidang lainnya [6]. Pada penelitian ini, t-SNE digunakan untuk menggambarkan data dengan dimensi tinggi ke dalam representasi yang lebih rendah sehingga dapat dengan mudah diinterpretasikan dan divisualisasikan.



Gambar 16. *2D t-SNE Visualization*



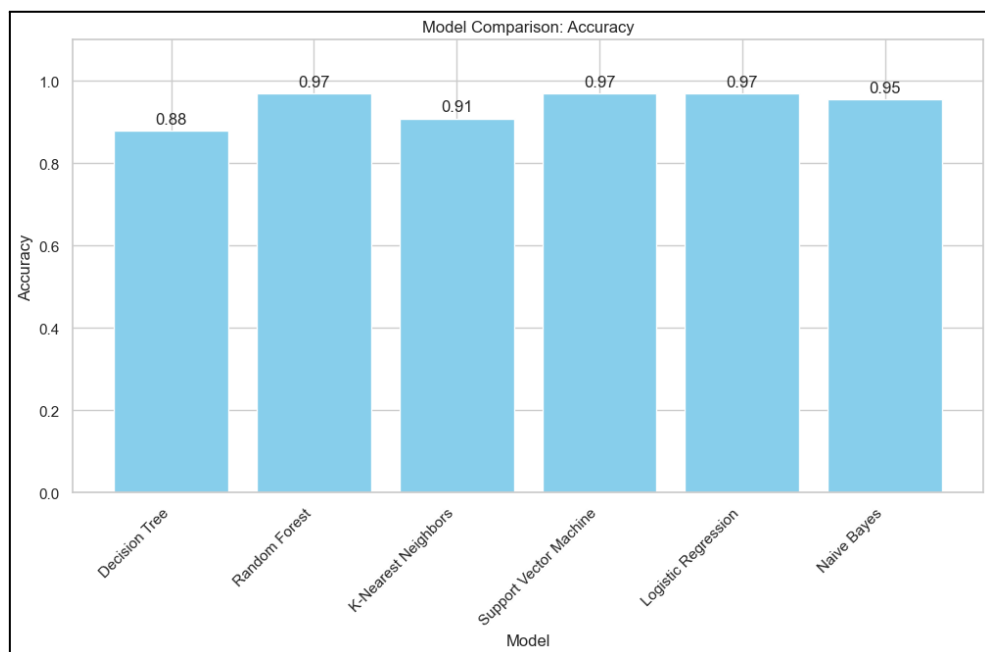
Gambar 17. *3D t-SNE Visualization*



Gambar 18. 3D Pie Chart

E. Visualizing Classification

Terdapat 6 model klasifikasi yang telah dilakukan pada penelitian ini, antara lain *Decision Tree*, *Random Forest*, *K-Nearest Neighbor*, *Support Vector Machine*, *Logistic Regression*, dan *Naïve Bayes*. Setelah dianalisis, hasil akurasi yang didapat untuk setiap model klasifikasi dapat dilihat pada **Gambar 19**.



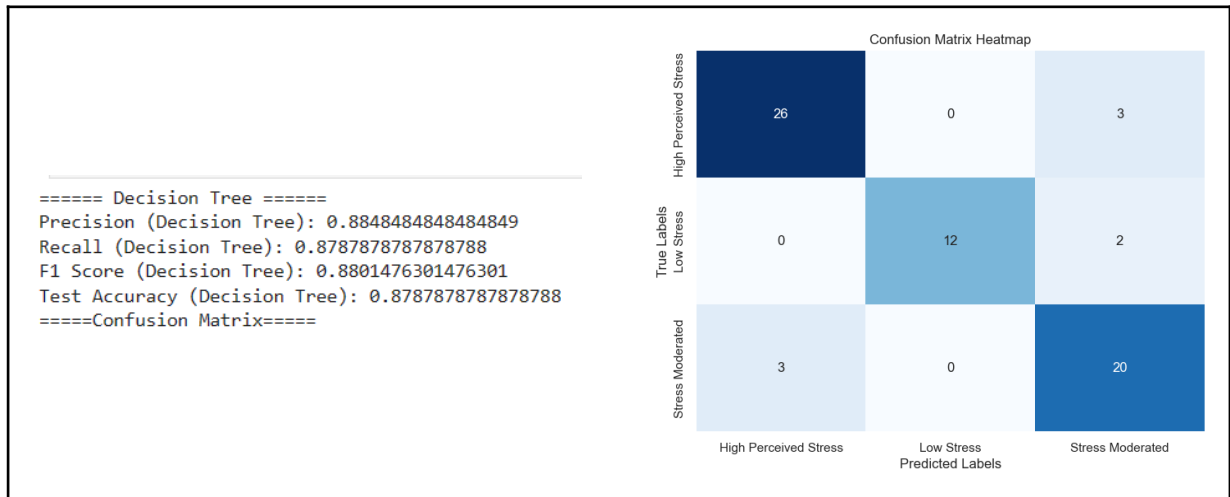
Gambar 19. Perbandingan akurasi model klasifikasi

Hasilnya menyatakan bahwa model klasifikasi yang memiliki nilai akurasi paling tinggi berada di nilai 0.97 dengan pembulatan. Terdapat 3 model klasifikasi yang memiliki nilai akurasi 0.97, yaitu *Random Forest*, *Support Vector Machine*, dan *Logistic Regression*. Model *Random Forest* dipilih sebagai

model evaluasi untuk digunakan pada proses prediksi tingkat stress. Setiap detail informasi model individu dapat dilihat sebagai berikut.

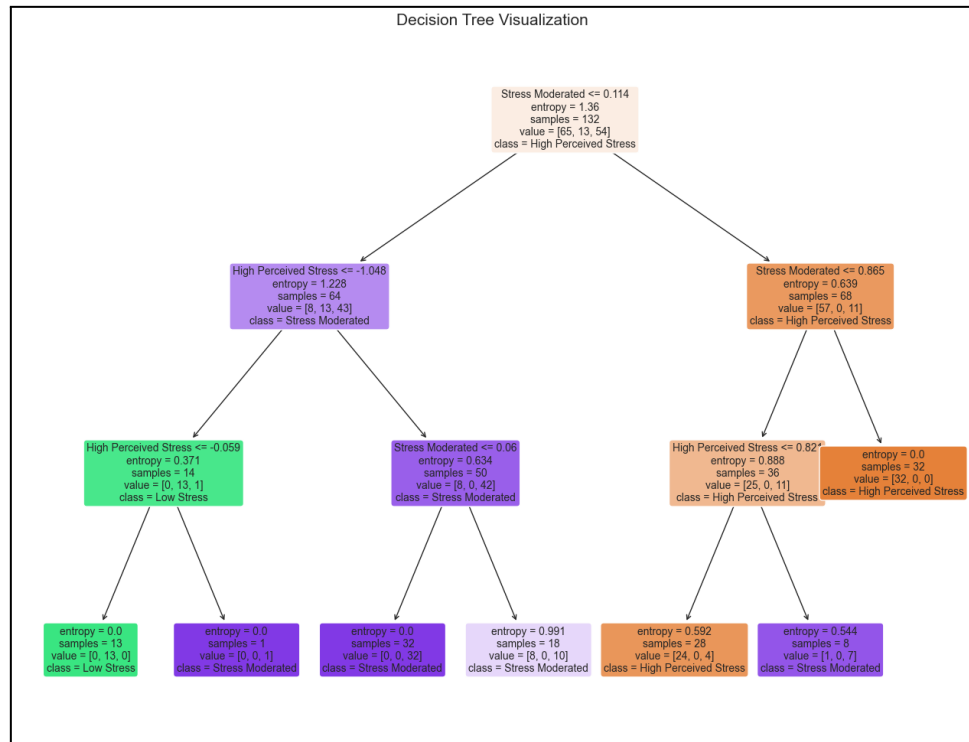
a. Decision Tree

Hasil evaluasi menggunakan model *Decision Tree* menghasilkan akurasi sebesar 0.88. Secara keseluruhan dapat dilihat pada **Gambar 20**.



Gambar 20. Evaluasi Model *Decision Tree*

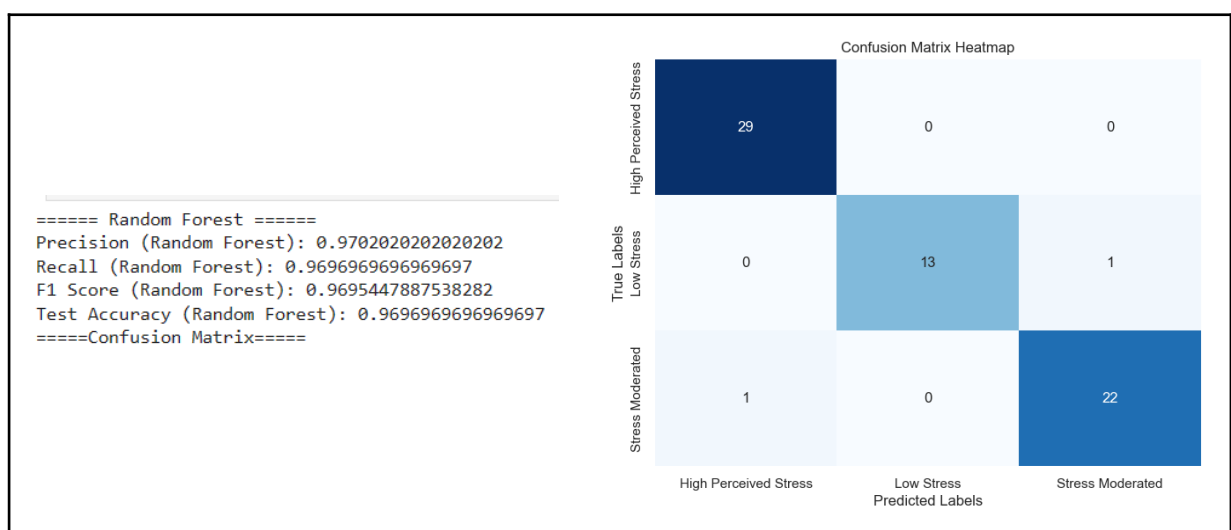
Model *Decision Tree* dapat divisualisasikan menjadi visualisasi pohon keputusan yang dapat dilihat pada Gambar. Visualisasi ini secara umum berguna untuk memahami bagaimana model *Decision Tree* membuat keputusan berdasarkan fitur input. Setiap kotak di dalam pohon merepresentasikan suatu keputusan atau pemisahan berdasarkan nilai dari salah satu fitur, dan cabang-cabangnya menunjukkan hasil keputusan tersebut. Daun-daun pohon menunjukkan prediksi akhir model.



Gambar 21. *Decision Tree Visualization*

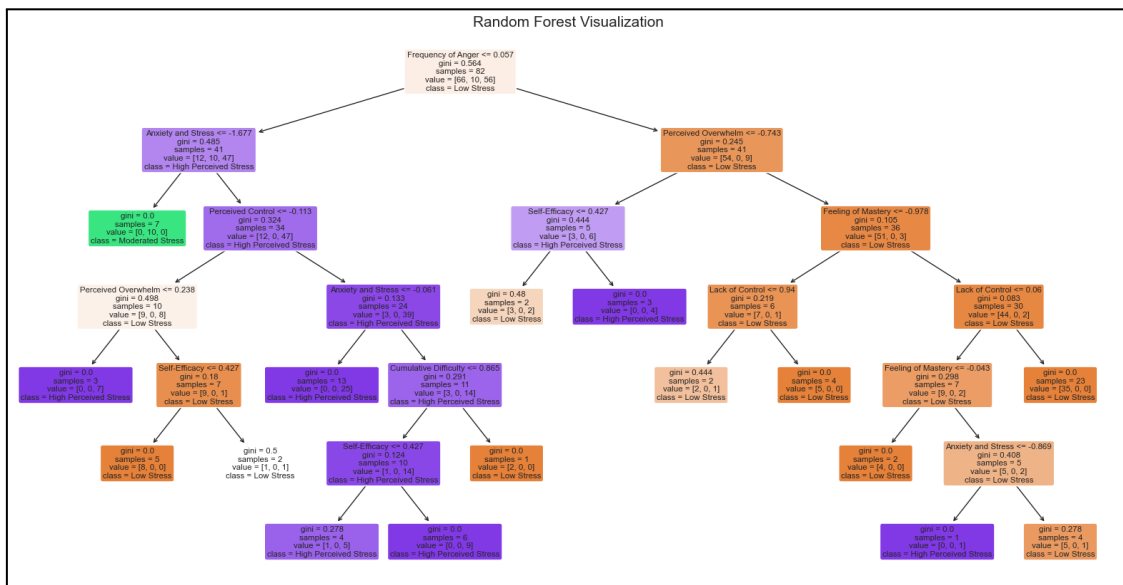
b. Random Forest

Hasil evaluasi menggunakan model *Random Forest* menghasilkan akurasi sebesar 0.97 dan digunakan sebagai model evaluasi. Secara keseluruhan dapat dilihat pada **Gambar 22**.



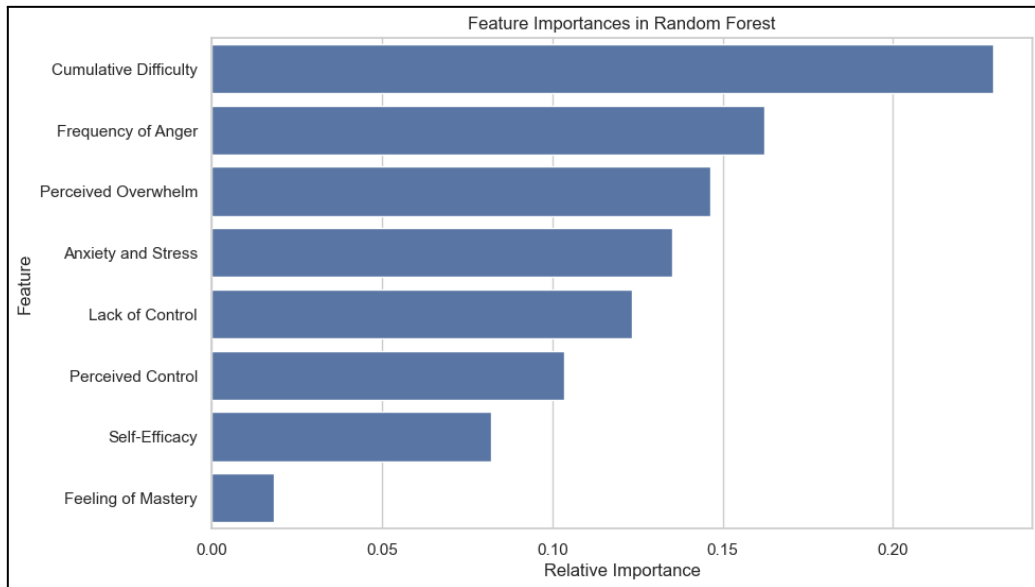
Gambar 22. *Evaluasi Model Decision Tree*

Model *Decision Tree* dapat divisualisasikan menjadi visualisasi pohon keputusan yang dapat dilihat pada Gambar. Random Forest merupakan ensemble dari banyak pohon keputusan, dan visualisasi ini dari salah satu pohon individu dalam ensemble. Visualisasi ini memberikan gambaran tentang cara pohon tersebut membuat keputusannya, beberapa aturan dan keputusan yang diambil oleh model ini untuk klasifikasi data dapat dilihat.



Gambar 23. *Random Forest Visualization*

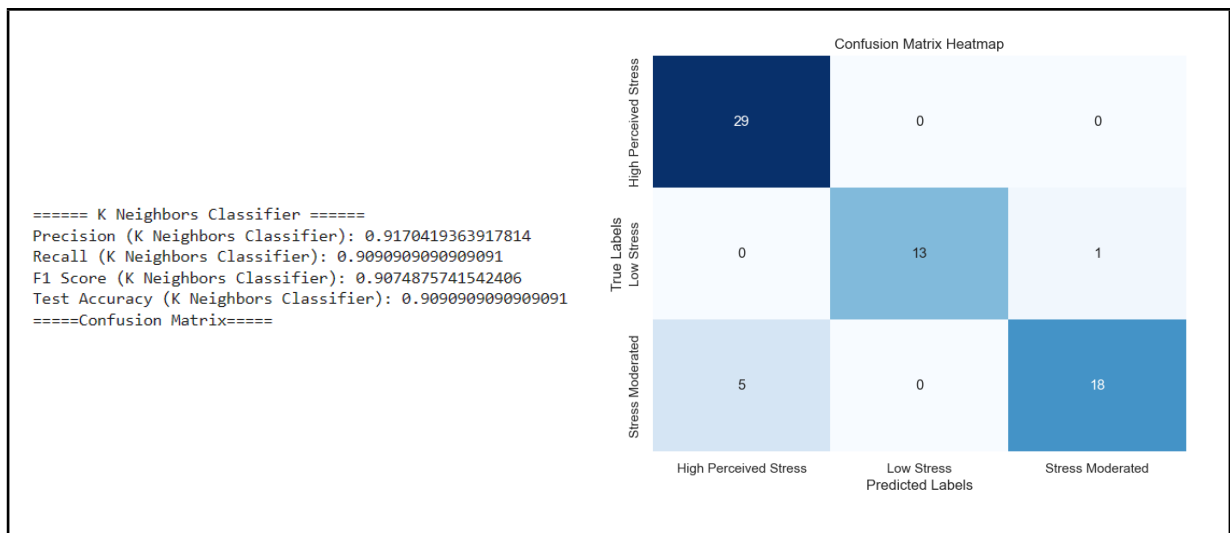
Selain itu, visualisasi barplot yang menunjukkan tingkat kepentingan relatif dari setiap fitur dalam model *Random Forest* dapat dilihat pada Gambar. Semakin panjang bar, semakin besar kontribusi fitur tersebut terhadap prediksi model. Visualisasi ini dapat membantu dalam memahami fitur-fitur yang paling berpengaruh dalam membuat keputusan oleh model *Random Forest*.



Gambar 24. *Random Forest - Feature Importance*

c. K-Nearest Neighbors

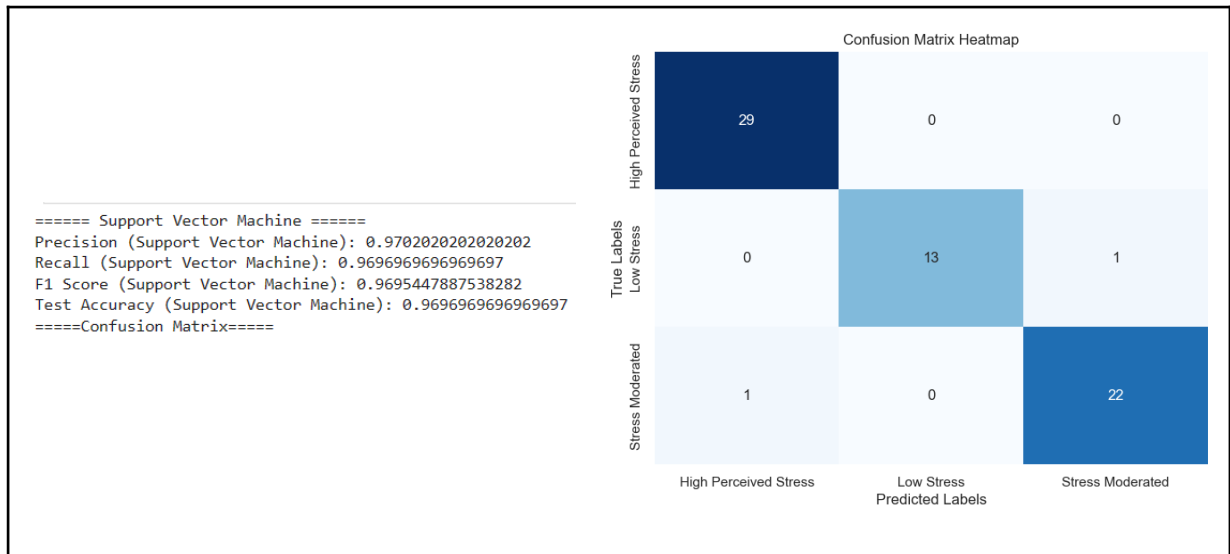
Hasil evaluasi menggunakan model *K-Nearest Neighbors* menghasilkan akurasi sebesar 0.91. Secara keseluruhan dapat dilihat pada **Gambar 25**.



Gambar 25. *Evaluasi Model K-Nearest Neighbors*

d. Support Vector Machine

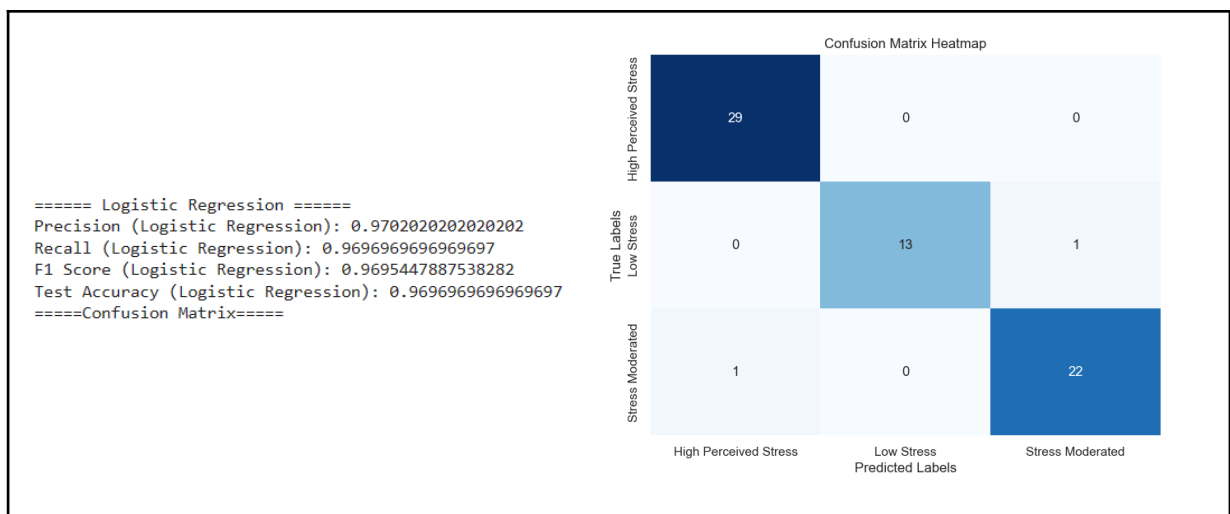
Hasil evaluasi menggunakan model *Support Vector Machine* menghasilkan akurasi sebesar 0.97. Secara keseluruhan dapat dilihat pada **Gambar 26**.



Gambar 26. Evaluasi Model *Support Vector Machine*

e. Logistic Regression

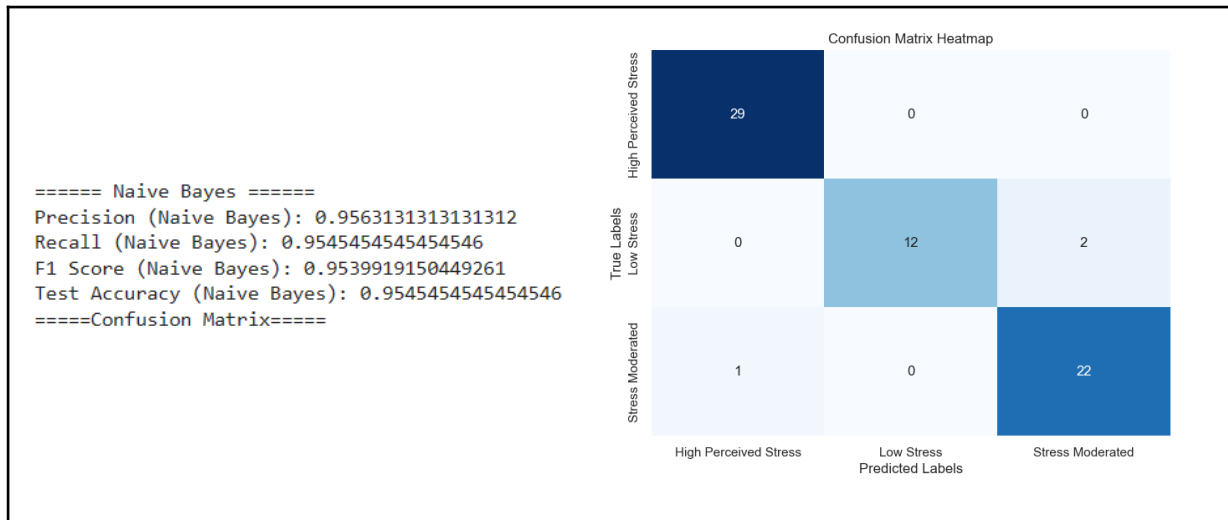
Hasil evaluasi menggunakan model *Logistic Regression* menghasilkan akurasi sebesar 0.97. Secara keseluruhan dapat dilihat pada **Gambar 27**.



Gambar 27. Evaluasi Model *Logistic Regression*

f. Naïve Bayes

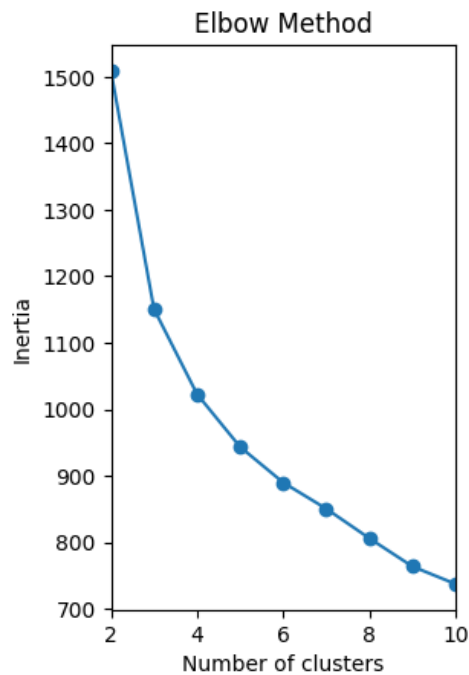
Hasil evaluasi menggunakan model *Naïve Bayes* menghasilkan akurasi sebesar 0.95. Secara keseluruhan dapat dilihat pada **Gambar 28**.



Gambar 28. Evaluasi Model *Naive Bayes*

F. Visualisasi kluster analisis (Clustering)

Clustering menggunakan K-Means. Penentuan jumlah kluster dari metode Elbow.



Gambar 29. *Elbow Method*

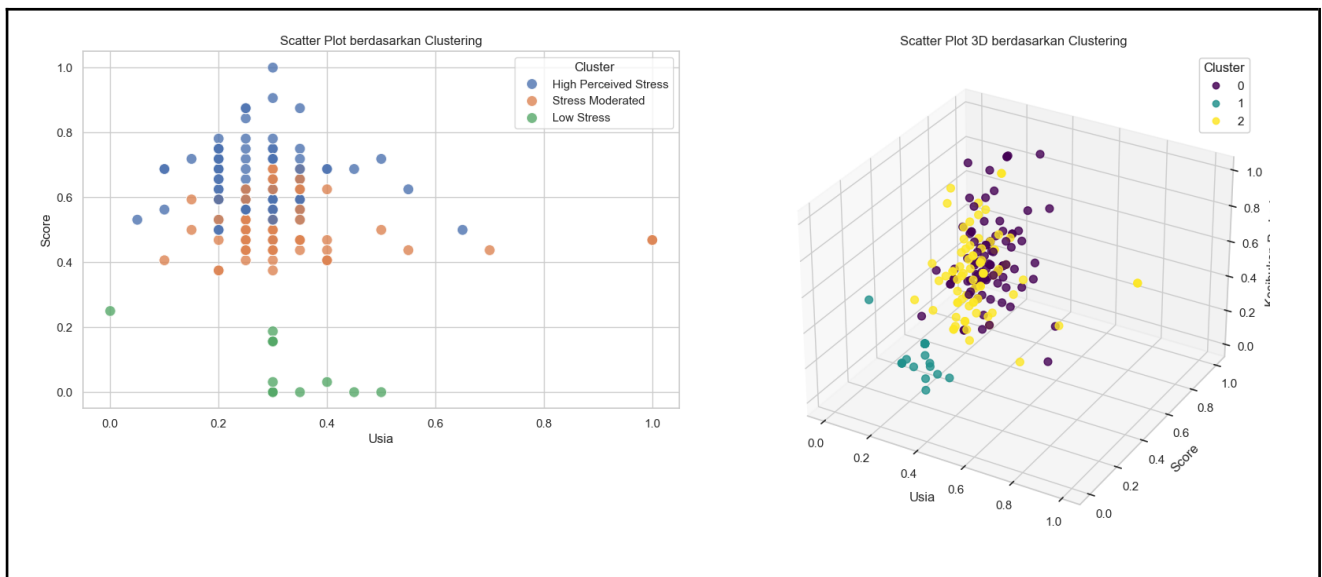
Berdasarkan visualisasi Elbow Method, didapat jumlah kluster yaitu 3 kluster dengan nilai *silhouette score* sebesar 0.28353458560010353. Hasil dari kluster ini lah yang menjadi variabel output yang menjadi label tingkatan stress. Label

tingkat stress ini mencakup ‘Low Stress’ dengan anggota 27 anggota, ‘Stress Moderated’ dengan anggota 77 anggota, dan ‘High Perceived Stress’ dengan anggota 94 anggota.

Jumlah Anggota Tiap Cluster:
High Perceived Stress: 94 anggota
Stress Moderated: 77 anggota
Low Stress: 27 anggota

Gambar 30. Anggota Tiap Cluster

Hasil dari klaster yang menjadi output tingkat stress dapat dilihat persebarannya di **Gambar 31**.



Gambar 31. Persebaran Data

G. Eksplorasi grafis data bisnis GUI

Hasil dari perhitungan *machine learning* yang kami lakukan diintegrasikan ke dalam aplikasi web menggunakan Streamlit untuk memberikan prediksi dan layanan berbasis data kepada pengguna. Streamlit merupakan sebuah framework aplikasi yang mempermudah pembuatan aplikasi web untuk pembelajaran mesin dengan cara yang sederhana [2]. Kerangka aplikasi ini sangat sesuai dengan pustaka-pustaka utama dalam bahasa pemrograman Python yang digunakan dalam pembelajaran mesin untuk keperluan visualisasi dan analisis data.

Pada penelitian ini, Streamlit digunakan untuk melakukan evaluasi tingkat stress berdasarkan model machine learning yang telah dipelajari sebelumnya dan disimpan menggunakan *library pickle*. Model yang digunakan adalah model *Random Forest*. Berikut adalah langkah-langkah bagaimana model machine learning dapat diintegrasikan ke dalam aplikasi web menggunakan Streamlit dengan bahasa pemrograman Python:

1. *Import library* yang akan digunakan seperti ***pickle, streamlit, pandas, numpy,*** dan ***sklearn***.
2. Persiapkan model *machine learning*. Berdasarkan hasil machine learning yang sudah dilakukan sebelumnya, simpan model ke dalam file .sav dengan menggunakan *library pickle*.
3. Load model dan objek penskalaan menggunakan *library pickle*.
4. Buat Antarmuka Web dengan Streamlit. Setiap elemen seperti judul, input, dan tombol dapat ditambahkan menggunakan perintah seperti ***st.title,*** ***st.number_input,*** ***st.selectbox,*** dan lain-lain.
5. Proses Input user dan prediksi. Ambil input dari user menggunakan elemen-elemen antarmuka web yang telah dibuat, lalu gunakan model *machine learning* untuk melakukan prediksi berdasarkan input pengguna. Total input dari user adalah 8 variabel yang dapat dilihat pada **Gambar 32**.
6. Menampilkan hasil dari prediksi. Hasil dari prediksi bisa ditampilkan dengan perintah ***st.success***. Selain itu, kami juga menambahkan beberapa kalimat motivasi yang dapat membantu pengguna dalam mengatasi masalah stres yang mereka miliki. Contoh dari output Streamlit dapat dilihat pada **Gambar 33**.

Stress Level Evaluation

Dalam sebulan terakhir, seberapa sering Anda kesal karena suatu hal terjadi secara tak terduga?

Tidak pernah



Dalam sebulan terakhir, seberapa sering Anda merasa tidak mampu mengendalikan hal-hal penting dalam hidup Anda?

Tidak pernah



Dalam sebulan terakhir, seberapa sering Anda merasa gugup dan stres?

Tidak pernah



Dalam sebulan terakhir, seberapa sering Anda merasa yakin dengan kemampuan Anda dalam menangani masalah pribadi?

Tidak pernah



Dalam sebulan terakhir, seberapa sering Anda merasa segalanya berjalan sesuai keinginan Anda?

Tidak pernah



Dalam sebulan terakhir, seberapa sering Anda menyadari bahwa Anda tidak mampu menyelesaikan semua hal yang harus Anda lakukan?

Tidak pernah



Dalam sebulan terakhir, seberapa sering Anda merasa berada di puncak segalanya?

Tidak pernah



Dalam sebulan terakhir, seberapa sering Anda merasakan kesulitan yang menumpuk sehingga Anda tidak mampu mengatasinya?

Tidak pernah



Ukur Tingkat Stress

Gambar 32. *GUI Streamlit - Input User*

Input Data:

	Frequency of Anger	Lack of Control	Anxiety and Stress	Self-Efficacy	Perceived Control	Perceived Overwhelm
0	4	4	4	1	5	3

Scaled Data:

	Frequency of Anger	Lack of Control	Anxiety and Stress	Self-Efficacy	Perceived Control	Perceived Overwhelm
0	0.4756	0.5001	0.3427	-2.0043	2.2237	

Prediction:

	Stress Level
0	High Perceived Stress

Stress Level Anda adalah Tinggi 😞

Penting untuk diakui bahwa tingkat stres yang tinggi dapat menjadi beban yang berat. Anda tidak sendirian dalam perasaan ini, dan ada sumber dukungan yang dapat diakses. Pertimbangkan untuk berbagi dengan teman, keluarga, atau mencari bantuan profesional. Saya di sini untuk mendengarkan dan memberikan dukungan sepanjang perjalanan ini. 🤝

Jangan ragu untuk mengirimkan pesan jika ada yang ingin dibicarakan atau ditanyakan. Kami di sini untuk membantu, termasuk jika Anda ingin berbagi pengalaman atau mencari dukungan untuk meredakan stres. Silakan kirimkan pertanyaan atau komentar Anda, dan kami akan berusaha memberikan respons secepat mungkin.

📞 Airlangga: +62 889-9645-9159

📞 Nathanael: +62 812-8777-4009

📞 Shabrina: +62 857-0630-6913

Gambar 33. *GUI Streamlit - Output*

KESIMPULAN DAN SARAN

Berdasarkan penelitian ini, dapat disimpulkan bahwa pemahaman mendalam terkait tingkat stres masyarakat dalam konteks dinamika modern memegang peranan krusial untuk merumuskan kebijakan kesehatan mental yang efektif. Pengukuran menggunakan *Perceived Stress Scale* (PSS) memberikan dimensi kuantitatif yang jelas terkait tingkat stres masyarakat, sementara pendekatan clustering memungkinkan identifikasi pola-pola tingkat stres yang mungkin tidak terlihat secara langsung.

Hasil penelitian menunjukkan bahwa melalui pendekatan clustering, ditemukan adanya tiga cluster sebagai label tingkat stres masyarakat. Analisis clustering memberikan pemahaman mendalam terkait pola-pola tingkat stres yang mungkin tidak terlihat secara langsung, memungkinkan pengelompokan masyarakat berdasarkan karakteristik stres yang signifikan. Pentingnya pendekatan ini semakin diperkuat dengan penggunaan model Random Forest sebagai alat evaluasi, yang mampu memberikan hasil prediksi yang cukup akurat terkait tingkat stres. Model ini memberikan keandalan dalam mengidentifikasi faktor-faktor pemicu stres yang dominan di setiap cluster.

Faktor-faktor pemicu stres yang dominan dapat diidentifikasi, memberikan wawasan berharga bagi perancangan intervensi yang sesuai dengan tingkat stres masyarakat. Keterlibatan faktor-faktor budaya dan sosial juga menjadi landasan penting dalam merancang kebijakan yang mendukung kesejahteraan masyarakat secara menyeluruh.

Penelitian lebih lanjut dapat dilakukan untuk mendalami pemahaman terhadap faktor-faktor pemicu stres yang kompleks, serta untuk memperluas cakupan geografis dan demografis penelitian. Peningkatan kolaborasi antara sektor kesehatan dan sektor lainnya, seperti pendidikan dan pekerjaan, juga dapat menjadi langkah strategis dalam mengurangi tingkat stres masyarakat.

DAFTAR PUSTAKA

- [1] Denovan, A., Dagnall, N., Dhingra, K., & Grogan, S. (2019). Evaluating the Perceived Stress Scale among UK university students: implications for stress measurement and management. *Studies in Higher Education*, 44(1), 120–133. <https://doi.org/10.1080/03075079.2017.1340445>
- [2] Jain, V., Kavitha, H., & Mohana Kumar, S. (2022). Credit Card Fraud Detection Web Application using Streamlit and Machine Learning. *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, 1–5. <https://doi.org/10.1109/ICDSIS55133.2022.9915901>
- [3] Purnami, C. (2020). Instrumen “Perceive Stress Scale” Online. Suatu Alternatif Alat Pengukur Tingkat Stress Secara Mudah Dan Cepat. Seminar Nasional Pengabdian Kepada Masyarakat UNDIP 2020, 1(1). Retrieved from <https://proceedings.undip.ac.id/index.php/semnasppm2019/article/view/119>
- [4] Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- [5] Suresh, M., Yuvaprasanth, R., Arun Ram Nathan, R. B., & Amarnath, K. (2020). Employees stress level assessment: a case of apparel industry. *IOP Conference Series: Materials Science and Engineering*, 954(1), 012018. <https://doi.org/10.1088/1757-899X/954/1/012018>
- [6] Yang, Z., Sun, Y., Li, D., Zhang, Z., & Xie, Y. (2020). T-Distributed Stochastic Neighbor Embedding with Gauss Initialization of Quantum Whale Optimization Algorithm. *2020 39th Chinese Control Conference (CCC)*, 3200–3205. <https://doi.org/10.23919/CCC50068.2020.9189639>