Data Mining A

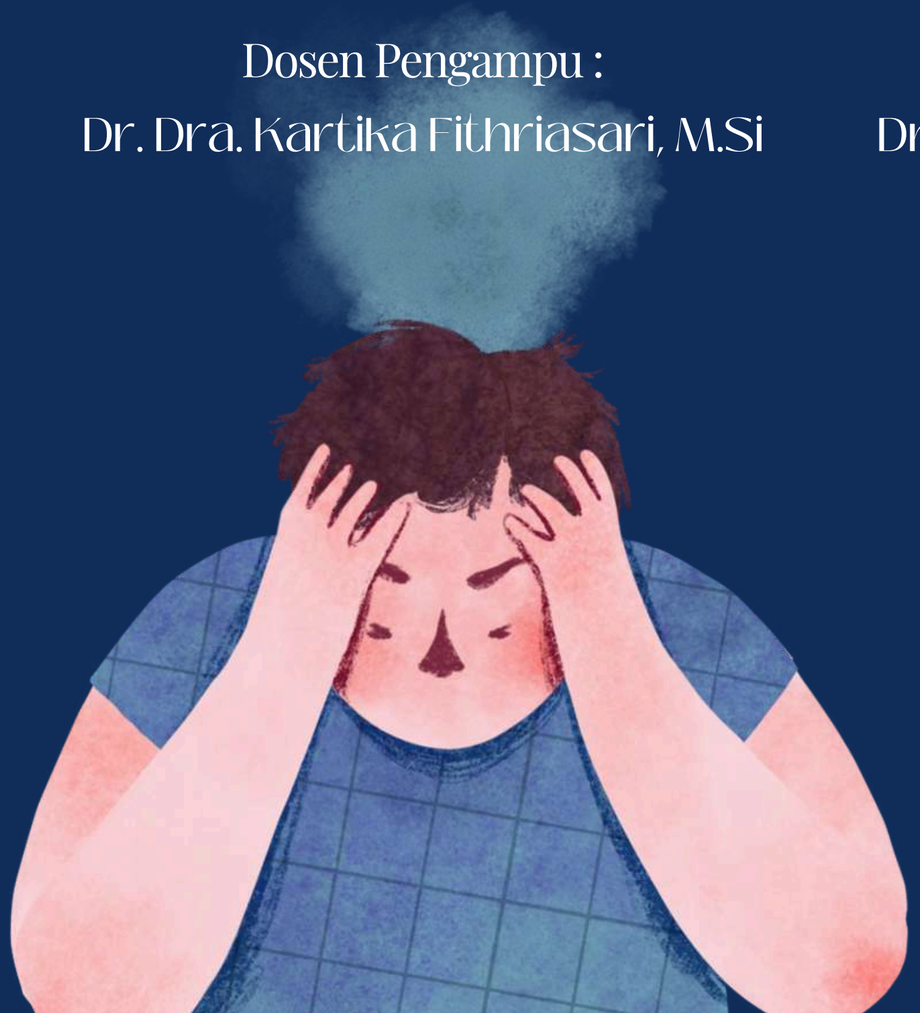
Psychological State Identification

Dosen Pengampu :

Dr. Irhamah, S.Si, M.Si

Dr. Dra. Kartika Fithriasari, M.Si

Dr. Santi Wulan Purnami



Kenalan Yuk!!

Brahmayudha Erlangga Putra

5003221084

OUTLINE

01 - ABOUT DATASET

Dalam era digital, kesehatan mental semakin menjadi perhatian. Biosensor, perangkat yang mengukur perubahan fisiologis, menawarkan pendekatan baru yang lebih objektif dalam mengidentifikasi kondisi psikologis. Dengan memantau parameter seperti detak jantung dan pernapasan secara real-time, biosensor dapat mendeteksi tanda-tanda awal gangguan mental dan memantau efektivitas terapi. Potensi biosensor dalam meningkatkan deteksi dini dan perawatan kesehatan mental sangat menjanjikan.

Oleh karena itu, diperlukan adanya analisis untuk mengklasifikasikan kondisi psikologis berdasarkan beberapa indikator Biosensor.

01 - ABOUT DATASET

Dataset berisi 20 kolom dengan variabel target yaitu ‘Psychological State’ yang memiliki 4 kategori

ID	Time	HRV (ms)	GSR (µS)	EEG Power Bands	Blood Pressure (mmHg)	Oxygen Saturation (%)	Heart Rate (BPM)	Ambient Noise (dB)	Cognitive Load	Mood State	Psychological State	Respiration Rate (BPM)	Skin Temp (°C)	Focus Duration (s)	Task Type	Age	Gender	Educational Level	Study Major
1	2024-01-01 00:00:00	33.039739	1.031806	[0.7583653347946298, 1.423247998317594, 0.6157...	114/79	98.433312	98	56.863054	Low	Anxious	Stressed	21	34.566484	27	Exam	22	Female	Postgraduate	Engineering
2	2024-01-01 00:00:01	49.914651	1.340983	[0.5520419333516282, 1.858065835142619, 0.3766...	113/86	98.944505	70	45.343430	Low	Neutral	Stressed	21	35.358593	282	Assignment	23	Male	Undergraduate	Arts
3	2024-01-01 00:00:02	67.894401	1.006014	[1.0261365005886114, 1.3504934190994182, 2.308...	124/74	95.990753	91	50.029264	High	Sad	Relaxed	17	34.359495	50	Group Discussion	18	Female	Postgraduate	Arts
4	2024-01-01 00:00:03	34.705373	0.849270	[1.6075723109471591, 1.6619672129812242, 0.344...	120/73	98.173643	95	60.802104	Low	Neutral	Anxious	12	34.802638	223	Exam	28	Female	High School	Engineering
5	2024-01-01 00:00:04	52.896549	0.879084	[1.055003922514022, 0.7643319894343756, 1.0745...	111/80	96.225051	65	40.696384	High	Anxious	Stressed	14	35.869862	201	Group Discussion	24	Female	High School	Engineering
6	2024-01-01 00:00:05	39.835588	1.301080	[1.9776042514080998, 2.110298943521322, 2.2395...	125/71	98.776385	74	58.512561	Low	Neutral	Stressed	19	34.162644	160	Group Discussion	22	Other	High School	Arts
7	2024-01-01 00:00:06	57.862620	1.031550	[1.8082576401092578, 2.4006225659581943, 1.005...	118/70	97.006389	77	46.605604	Moderate	Happy	Anxious	15	35.413763	289	Lecture	29	Female	High School	Arts
8	2024-01-01 00:00:07	37.384658	1.044636	[2.4999330944469107, 1.5340481389344187, 1.575...	114/73	97.496875	63	64.054257	Moderate	Neutral	Focused	14	35.470839	197	Assignment	28	Female	Undergraduate	Engineering
9	2024-01-01 00:00:08	71.586499	0.831950	[1.064794037262249, 2.4687465156651243, 1.1907...	118/78	96.951639	78	53.620557	Moderate	Sad	Focused	17	34.679122	280	Group Discussion	20	Other	Undergraduate	Science

Sumber : **Psychological State Identification Dataset**

02-PRE-PROCESSING

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 20 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   ID                        1000 non-null   int64
 1   Time                      1000 non-null   object
 2   HRV (ms)                  1000 non-null   float64
 3   GSR (µS)                  1000 non-null   float64
 4   EEG Power Bands           1000 non-null   object
 5   Blood Pressure (mmHg)     1000 non-null   object
 6   Oxygen Saturation (%)     1000 non-null   float64
 7   Heart Rate (BPM)          1000 non-null   int64
 8   Ambient Noise (dB)        1000 non-null   float64
 9   Cognitive Load            1000 non-null   object
 10  Mood State                1000 non-null   object
 11  Psychological State       1000 non-null   object
 12  Respiration Rate (BPM)    1000 non-null   int64
 13  Skin Temp (°C)            1000 non-null   float64
 14  Focus Duration (s)        1000 non-null   int64
 15  Task Type                 1000 non-null   object
 16  Age                       1000 non-null   int64
 17  Gender                    1000 non-null   object
 18  Educational Level         1000 non-null   object
 19  Study Major               1000 non-null   object
dtypes: float64(5), int64(5), object(10)
memory usage: 156.4+ KB
```

```
data.duplicated().sum()
```

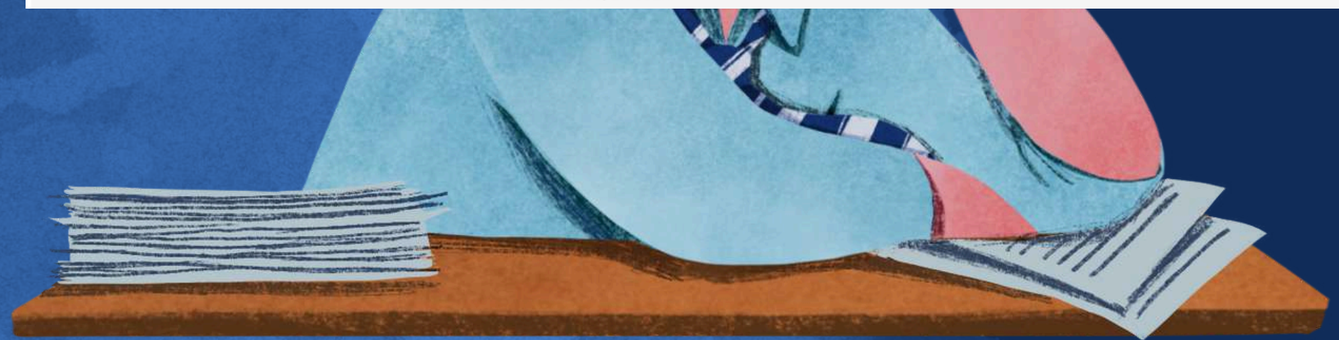
```
0
```

Berdasarkan informasi tersebut, maka tidak perlu dilakukan cleaning data karena tidak ada duplikasi data ataupun missing data. namun, demi keperluan analisis maka terdapat beberapa variabel yang dipisah ataupun dibuang

01 - ABOUT DATASET

Data Final :

	Psychological State	HRV (ms)	GSR (µS)	EEG Power Bands	EEG Delta	EEG Alpha	EEG Beta	Sistolik	Diastolik	Blood Pressure Category	...	Cognitive Load	Mood State	Respiration Rate (BPM)	Skin Temp (°C)	Focus Duration (s)	Task Type	Age	Gender	Educational Level	Study Major
0	Stressed	33.039739	1.031806	[0.7583653347946298, 1.423247998317594, 0.6157...	0.758365	1.423248	0.615770	114	79	Normal	...	Low	Anxious	21	34.566484	27	Exam	22	Female	Postgraduate	Engineering
1	Stressed	49.914651	1.340983	[0.5520419333516282, 1.858065835142619, 0.3766...	0.552042	1.858066	0.376605	113	86	Prehipertensi	...	Low	Neutral	21	35.358593	282	Assignment	23	Male	Undergraduate	Arts
2	Relaxed	67.894401	1.006014	[1.0261365005886114, 1.3504934190994182, 2.308...	1.026137	1.350493	2.308769	124	74	Prehipertensi	...	High	Sad	17	34.359495	50	Group Discussion	18	Female	Postgraduate	Arts
3	Anxious	34.705373	0.849270	[1.6075723109471591, 1.6619672129812242, 0.344...	1.607572	1.661967	0.344134	120	73	Prehipertensi	...	Low	Neutral	12	34.802638	223	Exam	28	Female	High School	Engineering
4	Stressed	52.896549	0.879084	[1.055003922514022, 0.7643319894343756, 1.0745...	1.055004	0.764332	1.074537	111	80	Prehipertensi	...	High	Anxious	14	35.869862	201	Group Discussion	24	Female	High School	Engineering
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	Stressed	73.056208	0.961572	[1.41709835038492, 0.5581367225529581, 2.05776...	1.417098	0.558137	2.057760	120	72	Prehipertensi	...	High	Neutral	22	34.710484	99	Assignment	24	Male	Postgraduate	Science
996	Focused	60.489220	1.167583	[0.4366569966688739, 0.6024761956899303, 1.601...	0.436657	0.602476	1.601388	110	86	Prehipertensi	...	Moderate	Happy	12	34.636933	191	Exam	19	Male	Undergraduate	Science
997	Relaxed	60.795800	0.996753	[2.4466973658165383, 0.4037428407944566, 0.437...	2.446697	0.403743	0.437608	127	83	Prehipertensi	...	Moderate	Happy	23	35.572915	103	Lecture	28	Male	Postgraduate	Science
998	Stressed	42.321587	0.681139	[2.232452864846464, 0.6589898799306273, 1.0954...	2.232453	0.658990	1.095468	126	73	Prehipertensi	...	Moderate	Happy	19	35.295968	241	Assignment	26	Male	Postgraduate	Engineering
999	Relaxed	49.005242	1.384472	[0.6375019674072194, 0.942131949788696, 2.4602...	0.637502	0.942132	2.460246	129	79	Prehipertensi	...	Low	Anxious	17	34.357262	94	Exam	23	Male	Undergraduate	Arts



03-SUMMARY STATISTICS

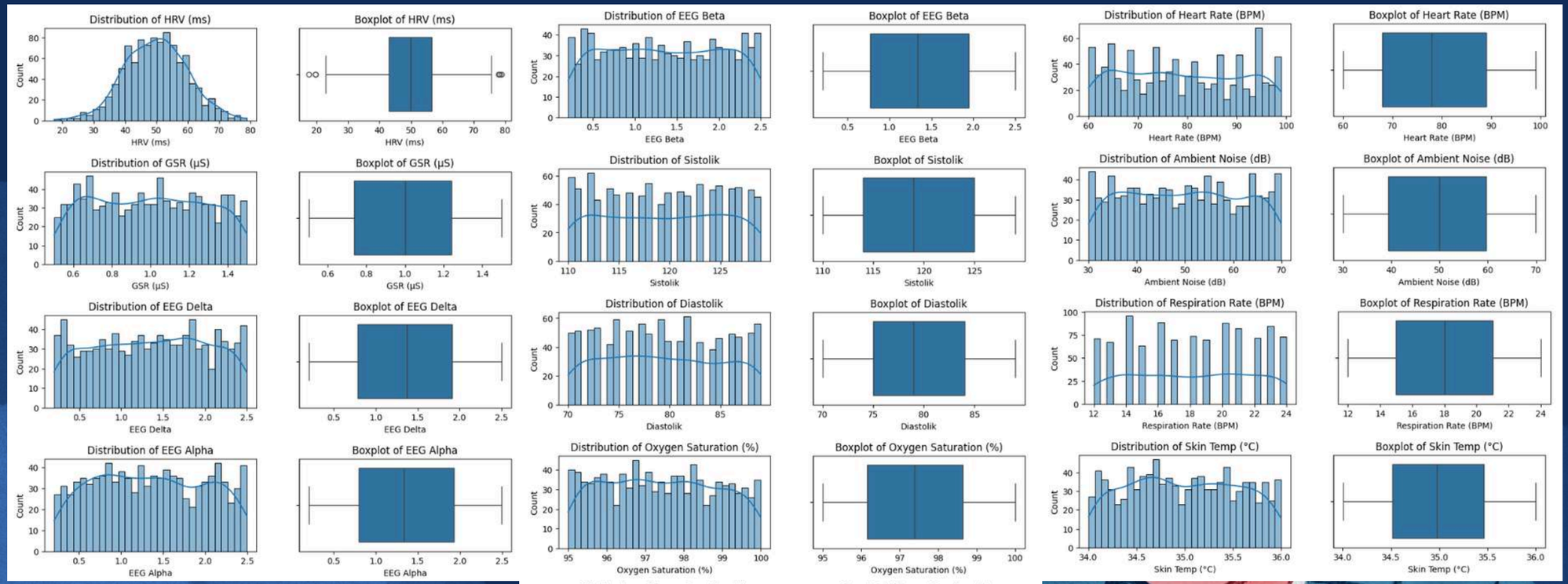
Berikut merupakan summary dari data numerik :

	count	mean	std	min	25%	50%	75%	max
HRV (ms)	1000.000000	49.920454	9.863658	17.405557	43.019097	49.982476	56.538365	78.841149
GSR (µS)	1000.000000	0.995749	0.285758	0.501567	0.736509	0.999965	1.239964	1.499563
EEG Delta	1000.000000	1.358240	0.670225	0.200521	0.787116	1.375709	1.909444	2.499933
EEG Alpha	1000.000000	1.350936	0.651778	0.203014	0.802376	1.336313	1.929745	2.494462
EEG Beta	1000.000000	1.350890	0.679453	0.202257	0.768054	1.336181	1.950509	2.499233
Sistolik	1000.000000	119.376000	5.854648	110.000000	114.000000	119.000000	125.000000	129.000000
Diastolik	1000.000000	79.382000	5.788401	70.000000	75.000000	79.000000	84.000000	89.000000
Oxygen Saturation (%)	1000.000000	97.419577	1.442865	95.001452	96.154447	97.395945	98.643364	99.997277
Heart Rate (BPM)	1000.000000	78.829000	11.818097	60.000000	68.000000	78.000000	89.000000	99.000000
Ambient Noise (dB)	1000.000000	49.825774	11.745028	30.017619	39.462492	50.012015	59.690683	69.902683
Respiration Rate (BPM)	1000.000000	18.061000	3.719558	12.000000	15.000000	18.000000	21.000000	24.000000
Skin Temp (°C)	1000.000000	34.988254	0.567596	34.001039	34.512701	34.974199	35.459442	35.997964
Focus Duration (s)	1000.000000	154.499000	84.098922	10.000000	83.750000	156.000000	227.000000	299.000000
Age	1000.000000	23.477000	3.420698	18.000000	21.000000	23.000000	27.000000	29.000000

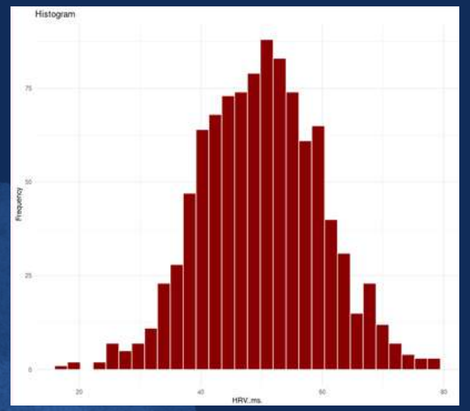
Berikut merupakan summary dari data kategorik :

	Psychological State	Blood Pressure Category	Cognitive Load	Mood State	Task Type	Gender	Educational Level	Study Major
count	1000	1000	1000	1000	1000	1000	1000	1000
unique	4	2	3	4	4	3	3	3
top	Stressed	Prehipertensi	High	Happy	Group Discussion	Other	Undergraduate	Arts
freq	263	726	351	268	259	343	338	337

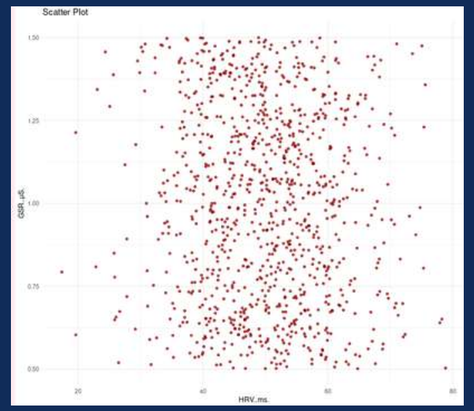
03-VISUALIZATION



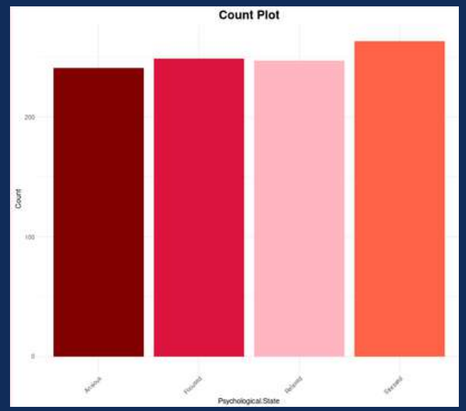
03-VISUALIZATION (R-SHINY)



Histogram

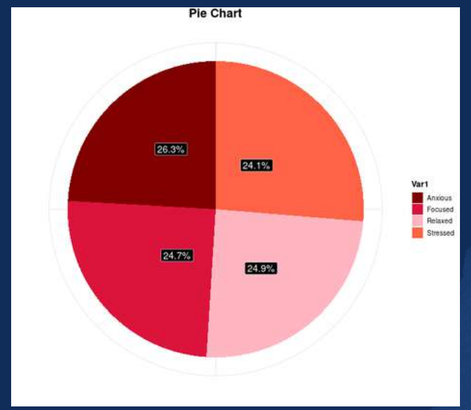


Scatter Plot



Count Plot

Count Plot



Pie Chart

04- FEATURE SELECTION

Mutual Information & Recursive Feature Elimination

Dengan menggunakan kedua metode feature selection didapatkan 4 fitur terbaik yang paling signifikan dan relevan dalam mempengaruhi hasil prediksi target. Fitur-fitur tersebut adalah :

- GSR (µS)
- EEG Delta
- EEG Beta
- Cognitive Load

05 - CLASSIFICATION & TRANING-TESTING

Metode Analisis Klasifikasi

Metode analisis kalisifkasi yang digunakan untuk klasifikasi kondisi psikologi yaitu :

- Decision Tree
- Logistic Regression
- Gaussian Naive Bayes

Analisis dilakukan mengguanakn traning-testing yaitu:

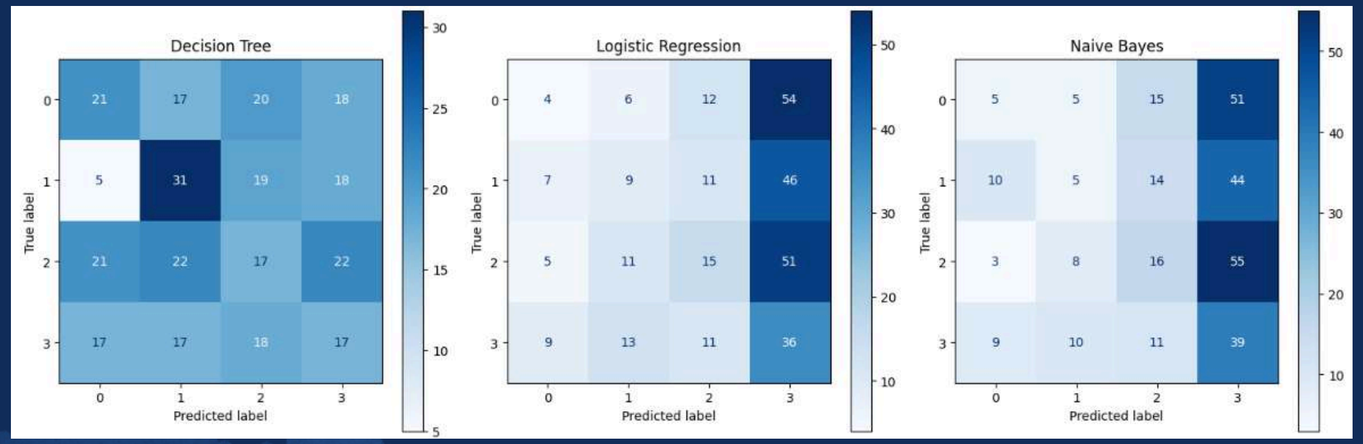
- Repeated Holdout (Stratified)
- K-Fold Cross Validation (CV)

Repeated Holdout

- Repeated = 10
- Train = 667
- Test = 333

05 - CLASSIFICATION & TRANING-TESTING

Evaluasi Model

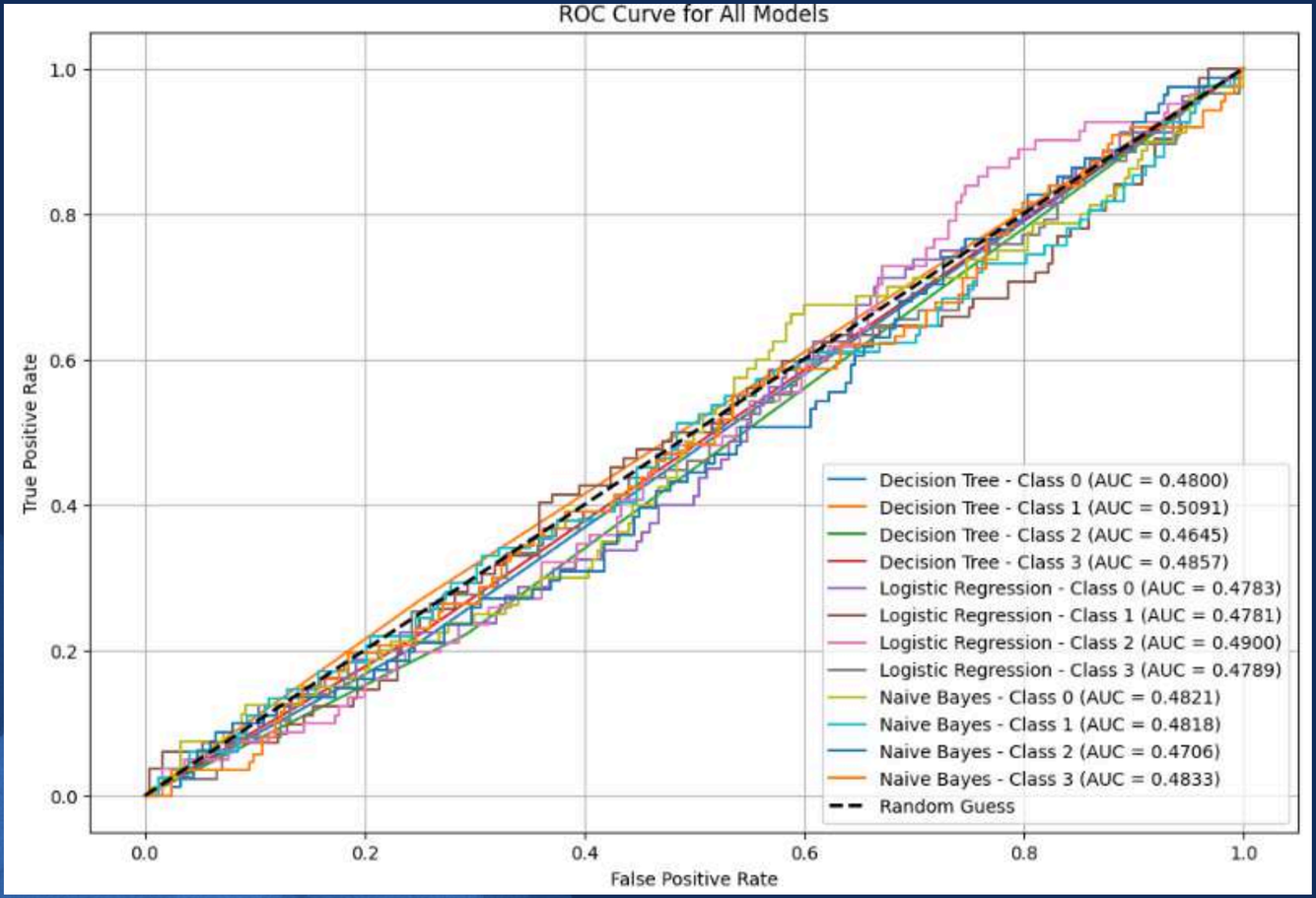


	Model	Accuracy	Sensitivity	Specificity	AUC
0	Decision Tree	0.227273	0.227226	0.742478	0.484852
1	Logistic Regression	0.251515	0.247477	0.748914	0.481351
2	Naive Bayes	0.239394	0.234139	0.744654	0.479452

- **Akurasi** : Ketiga variabel mengklasifikasikan 22,7%-25,1$% kondisi psikologis dengan benar secara keseluruhan.
- **Sensitivitas** : Ketiga model berhasil mendeteksi 22,7%-24,7% kondisi psikologis secara True Positif dengan benar.
- **Spesifisitas** : Ketiga model mampu mengidentifikasi 74,2%-74,8% kondisi psikologis secara True Negatif dengan benar.
- **AUC** : Ketiga model memiliki nilai AUC di angka 47,9%-48,4%. Hal ini menunjukkan model masih belum sepenuhnya baik untuk mengklasifikasikan kondisi psikologis.

05 - CLASSIFICATION & TRANING-TESTING

Kurva ROC



Berdasarkan Kurva ROC dan Nilai AUC, diketahui bahwa metode klasifikasi terbaik untuk memprediksi kondisi psikologis seseorang adalah Decision Tree dengan nilai AUC sebesar 48,4%.

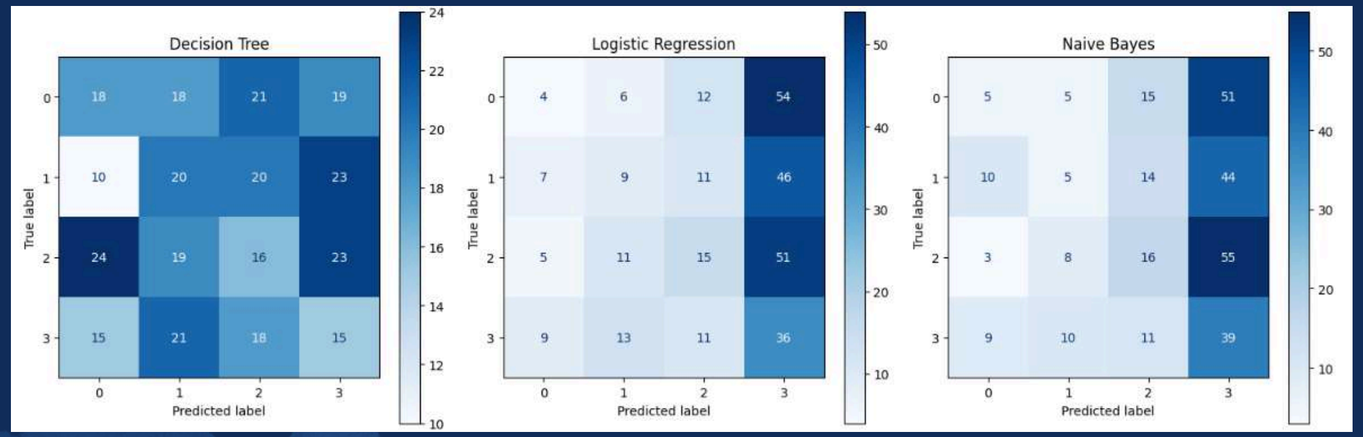
05 - CLASSIFICATION & TRANING-TESTING

K-Fold CV

K 10
1000 data per kelompok

05 - CLASSIFICATION & TRANING-TESTING

Evaluasi Model

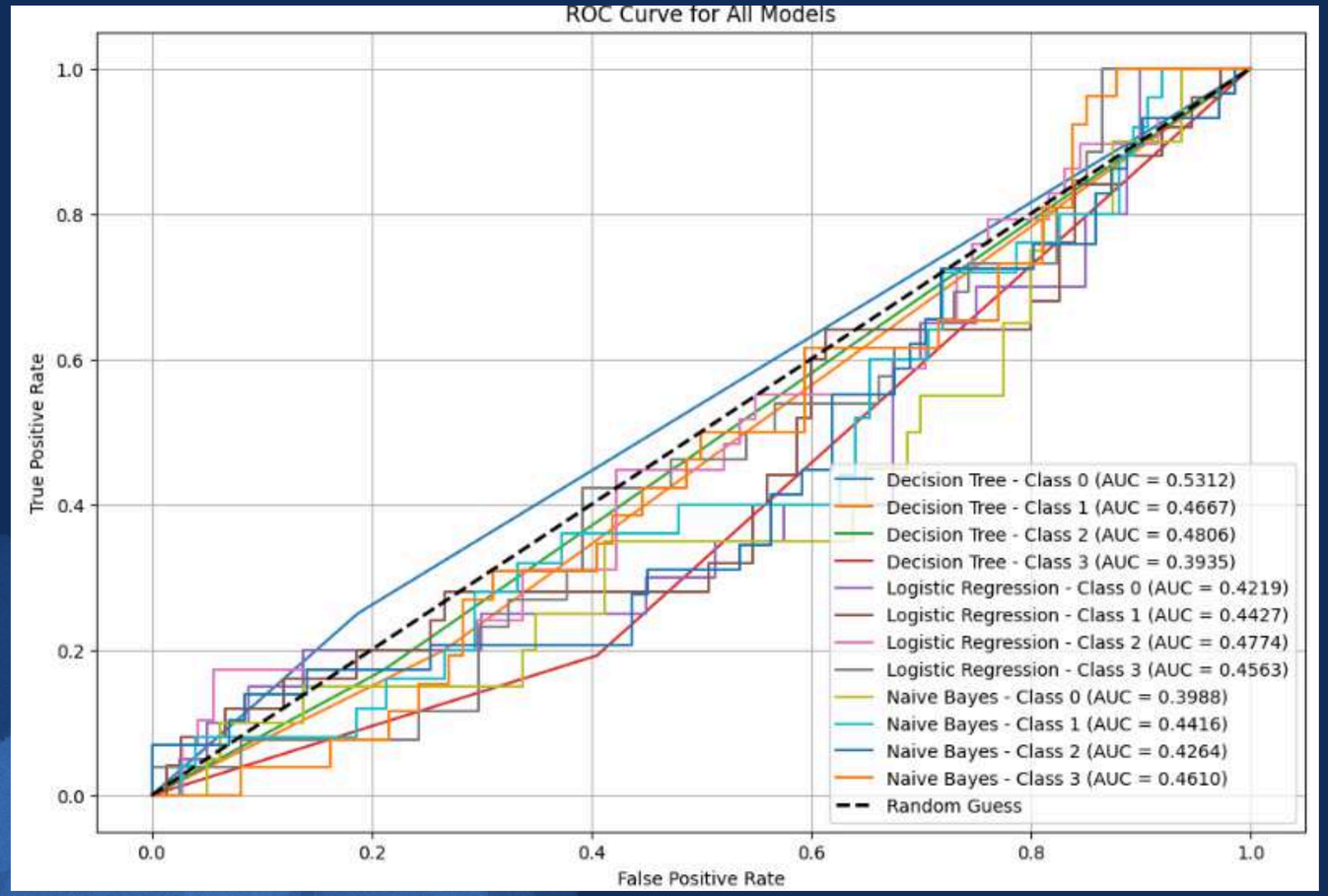


	Model	Accuracy	Sensitivity	Specificity	AUC
0	Decision Tree	0.244000	0.248461	0.732290	0.498648
1	Logistic Regression	0.238000	0.243336	0.754984	0.467289
2	Naive Bayes	0.234000	0.236879	0.744506	0.456281

- **Akurasi** : Ketiga model mengklasifikasikan 23,4%-24,4$% kondisi psikologis dengan benar secara keseluruhan.
- **Sensitivitas** : Ketiga model berhasil mendeteksi 23,6%-24,8% kondisi psikologis secara True Positif dengan benar.
- **Spesifisitas** : Ketiga model mampu mengidentifikasi 73,2%-75,4% kondisi psikologis secara True Negatif dengan benar.
- **AUC** : Ketiga model memiliki nilai AUC di angka 45,6%-49,8%. Hal ini menunjukkan model masih belum sepenuhnya baik untuk mengklasifikasikan kondisi psikologis.

05 - CLASSIFICATION & TRANING-TESTING

Kurva ROC



Berdasarkan Kurva ROC dan Nilai AUC, diketahui bahwa metode klasifikasi terbaik untuk memprediksi kondisi psikologis seseorang adalah Decision Tree dengan nilai AUC sebesar 49,8%.

Thanks