M.Asjaun

1103210181

Uts

Informasi Dataset:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 4000 entries, 0 to 3999

Data columns (total 91 columns):

#	Column	Non-Null Count	Dtype
0	Year	4000 non-null	int64
1	x1	4000 non-null	float64
2	x2	4000 non-null	float64
3	x3	4000 non-null	float64
4	x4	4000 non-null	float64
5	x5	4000 non-null	float64
6	x6	4000 non-null	float64
7	x7	4000 non-null	float64
8	x8	4000 non-null	float64
9	x9	4000 non-null	float64
10	x10	4000 non-null	float64
11	x11	4000 non-null	float64
12	x12	4000 non-null	float64
13	x13	4000 non-null	float64
14	x14	4000 non-null	float64
15	x15	4000 non-null	float64
16	x16	4000 non-null	float64
17	x17	4000 non-null	float64
18	x18	4000 non-null	float64
19	x19	4000 non-null	float64
20	x20	4000 non-null	float64
21	x21	4000 non-null	float64

22	x22	4000 non-null	float64
23	x23	4000 non-null	float64
24	x24	4000 non-null	float64
25	x25	4000 non-null	float64
26	x26	4000 non-null	float64
27	x27	4000 non-null	float64
28	x28	4000 non-null	float64
29	x29	4000 non-null	float64
30	x30	4000 non-null	float64
31	x31	4000 non-null	float64
32	x32	4000 non-null	float64
33	x33	4000 non-null	float64
34	x34	4000 non-null	float64
35	x35	4000 non-null	float64
36	x36	4000 non-null	float64
37	x37	4000 non-null	float64
38	x38	4000 non-null	float64
39	x39	4000 non-null	float64
40	x40	4000 non-null	float64
41	x41	4000 non-null	float64
42	x42	4000 non-null	float64
43	x43	4000 non-null	float64
44	x44	4000 non-null	float64
45	x45	4000 non-null	float64
46	x46	4000 non-null	float64
47	x47	4000 non-null	float64
48	x48	4000 non-null	float64
49	x49	4000 non-null	float64
50	x50	4000 non-null	float64
51	x51	4000 non-null	float64
52	x52	4000 non-null	float64
53	x53	4000 non-null	float64

54	x54	4000 non-null	float64
55	x55	4000 non-null	float64
56	x56	4000 non-null	float64
57	x57	4000 non-null	float64
58	x58	4000 non-null	float64
59	x59	4000 non-null	float64
60	x60	4000 non-null	float64
61	x61	4000 non-null	float64
62	x62	4000 non-null	float64
63	x63	4000 non-null	float64
64	x64	4000 non-null	float64
65	x65	4000 non-null	float64
66	x66	4000 non-null	float64
67	x67	4000 non-null	float64
68	x68	4000 non-null	float64
69	x69	4000 non-null	float64
70	x70	4000 non-null	float64
71	x71	4000 non-null	float64
72	x72	4000 non-null	float64
73	x73	4000 non-null	float64
74	x74	4000 non-null	float64
75	x75	4000 non-null	float64
76	x76	4000 non-null	float64
77	x77	4000 non-null	float64
78	x78	4000 non-null	float64
79	x79	4000 non-null	float64
80	x80	4000 non-null	float64
81	x81	4000 non-null	float64
82	x82	4000 non-null	float64
83	x83	4000 non-null	float64
84	x84	4000 non-null	float64
85	x85	4000 non-null	float64

86	x86	4000 non-null float64
87	x87	4000 non-null float64
88	x88	4000 non-null float64
89	x89	4000 non-null float64
90	x90	4000 non-null float64

dtypes: float64(90), int64(1)

memory usage: 2.8 MB

None

Ringkasan Statistik:

	Year	x1	x2	x3	x4 \
count	4000.000000	4000.000000	4000.000000	4000.000000	4000.000000
mean	1998.435000	43.311786	1.012153	8.463853	0.733843
std	11.034773	6.203113	51.931772	34.550414	15.871093
min	1927.000000	4.836880	-287.287360	-181.649920	-92.331640
25%	1994.000000	39.850093	-27.499023	-11.580035	-8.685952
50%	2002.000000	44.244950	7.177550	10.800570	-0.780245
75%	2006.000000	47.776155	36.072565	28.977497	8.317600
max	2010.000000	56.111220	177.001290	153.307500	109.984410

	x5	x6	x7	x8	x9 ... \
count	4000.000000	4000.000000	4000.000000	4000.000000	4000.000000 ...
mean	-6.717566	-9.437315	-2.249602	-1.893381	3.815266 ...
std	22.870201	12.694786	14.559507	7.942464	10.587308 ...
min	-102.797980	-70.450570	-66.829680	-39.386550	-74.424430 ...
25%	-20.770273	-18.333112	-10.703190	-6.561155	-2.065132 ...
50%	-6.277740	-11.006495	-1.860480	-1.821385	4.063490 ...
75%	7.200255	-2.281867	6.619840	2.520313	9.958815 ...
max	168.918870	57.099910	131.396350	52.576250	55.831130 ...

	x81	x82	x83	x84	x85 \
count	4000.000000	4000.000000	4000.000000	4000.000000	4000.000000

mean	16.298369	-75.223614	39.250547	34.652395	0.662173
std	35.048044	176.846479	117.832238	95.055468	16.882065
min	-196.289070	-1783.416650	-701.391740	-1147.976630	-200.135010
25%	-1.470347	-147.562252	-23.325915	-5.634960	-6.246187
50%	9.131615	-50.506850	28.999400	32.555595	0.881840
75%	26.071280	13.461400	88.818755	76.210900	8.686333
max	644.414220	1404.511780	921.008720	530.860360	133.924920

	x86	x87	x88	x89	x90
count	4000.000000	4000.000000	4000.000000	4000.000000	4000.000000
mean	18.797449	-26.717331	4.212853	20.868842	1.483193
std	117.535274	175.062616	13.861867	213.872855	24.067136
min	-1308.055150	-2179.274330	-70.480920	-7458.378150	-119.192560
25%	-30.921385	-102.911478	-2.739300	-57.671635	-8.756602
50%	15.873315	-19.234530	3.129655	11.952900	0.171950
75%	70.179970	57.769723	9.950910	84.542975	9.601770
max	748.164480	981.440830	243.004150	2042.531890	600.766240

[8 rows x 91 columns]

Pemeriksaan Nilai Null:

Year	0
x1	0
x2	0
x3	0
x4	0
..	
x86	0
x87	0
x88	0
x89	0
x90	0

Length: 91, dtype: int64

Distribusi x1

Distribusi x2

Distribusi x89

Matriks Korelasi (10 Kolom Pertama)
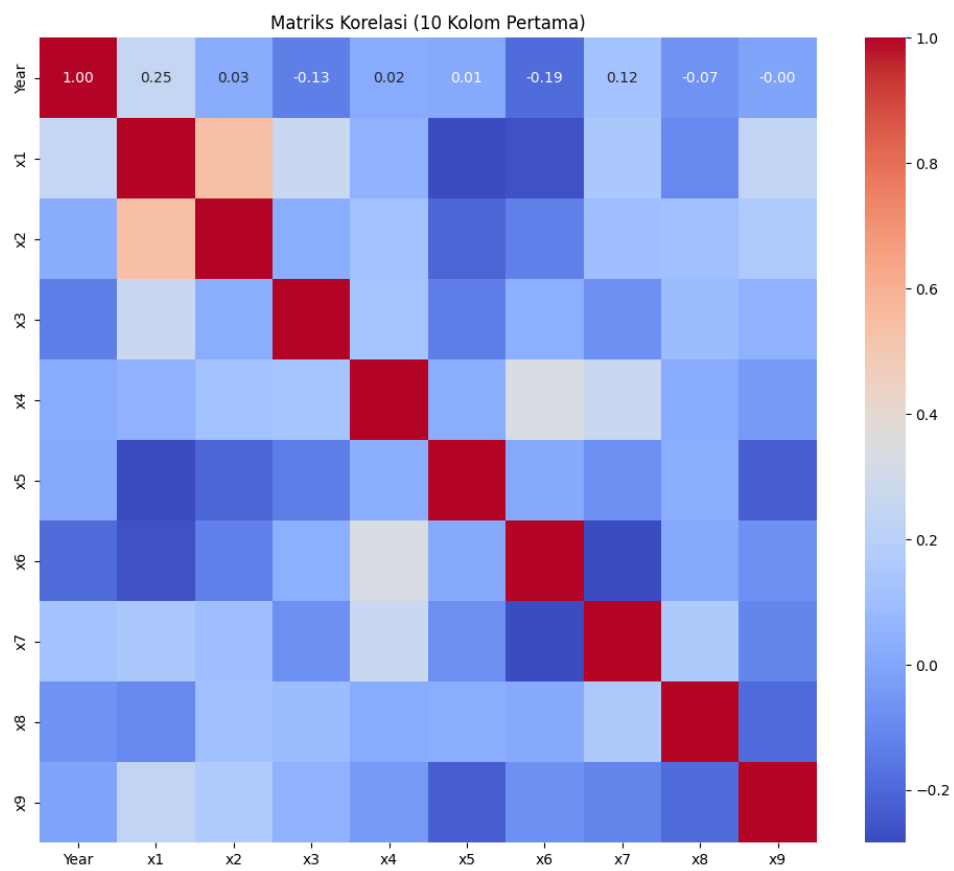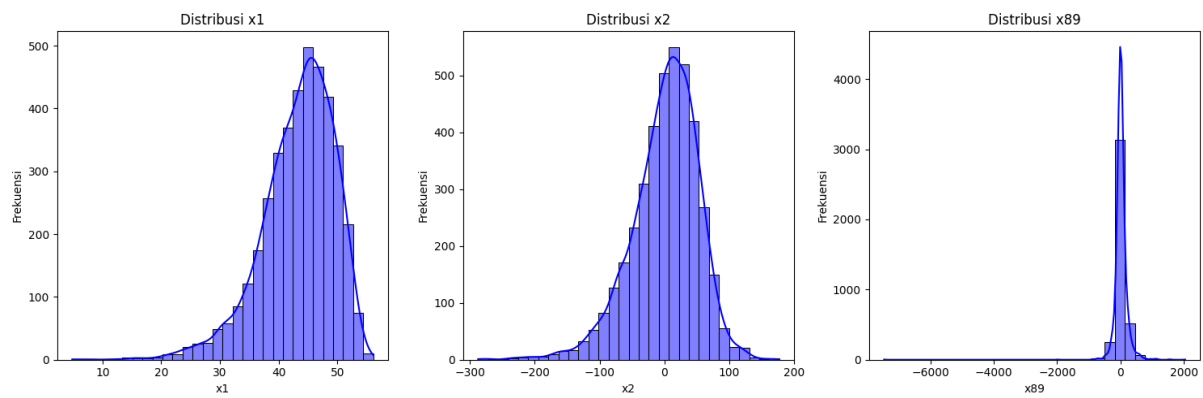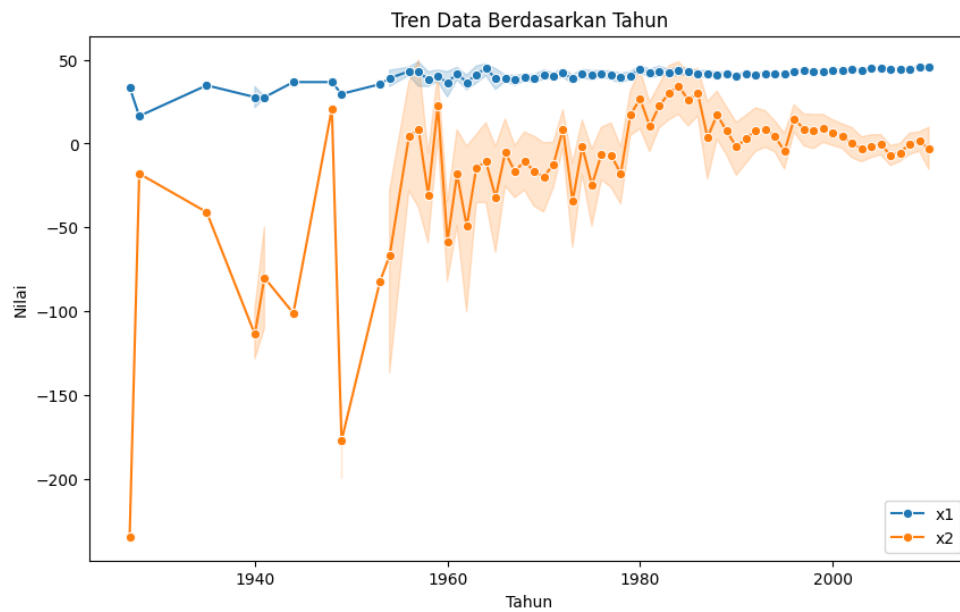
Hasil Evaluasi Model:

	Model	Best Parameters	MSE \
0	LinearRegression	{'poly__degree': 1}	95.419414
1	Ridge	{'poly__degree': 1, 'regressor__alpha': 100}	94.735538
2	Lasso	{'poly__degree': 1, 'regressor__alpha': 0.1}	94.464910

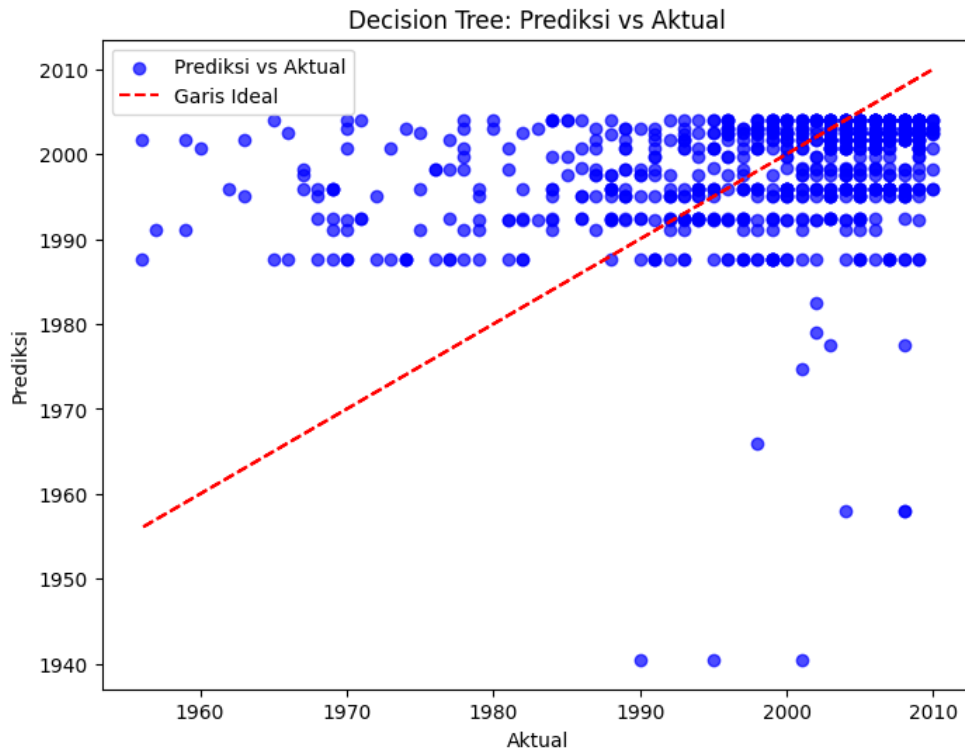|	R2
0	0.195256
1	0.201023
2	0.203306
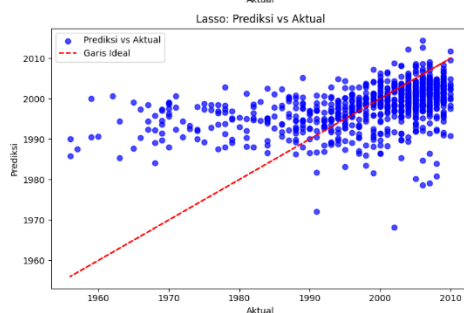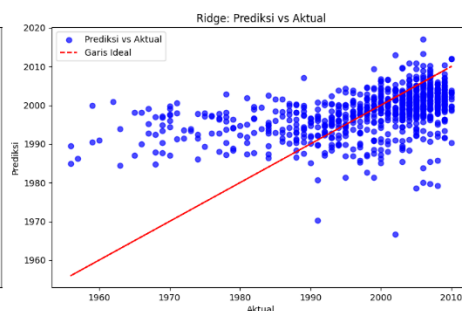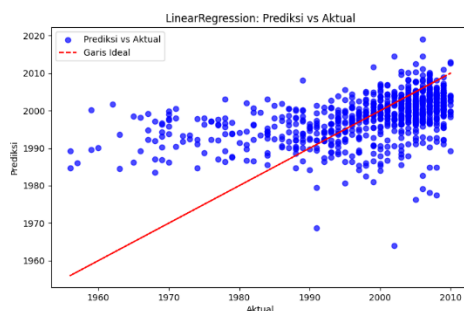
Laporan Evaluasi:

Best Parameters: {'regressor__max_depth': 5, 'regressor__min_samples_leaf': 2, 'regressor__min_samples_split': 10}

MSE: 128.81943893636526

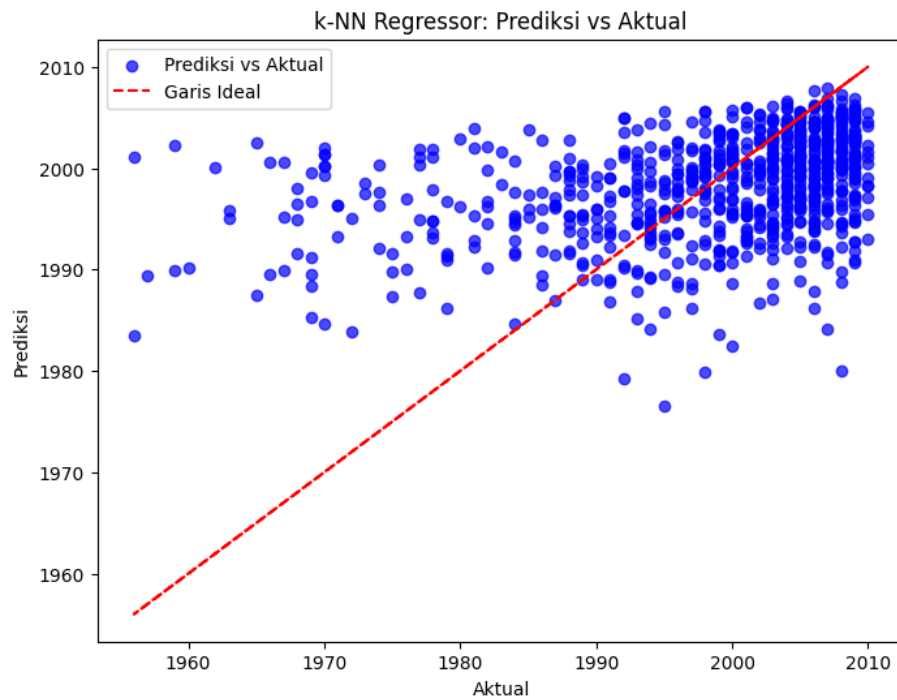R2: -0.08643207093942529

Decision Tree: Prediksi vs Aktual

Laporan Evaluasi:

Best Parameters: {'regressor__n_neighbors': 9, 'regressor__p': 1, 'regressor__weights': 'distance'}

MSE: 104.07453402408778

R2: 0.12226048750530494
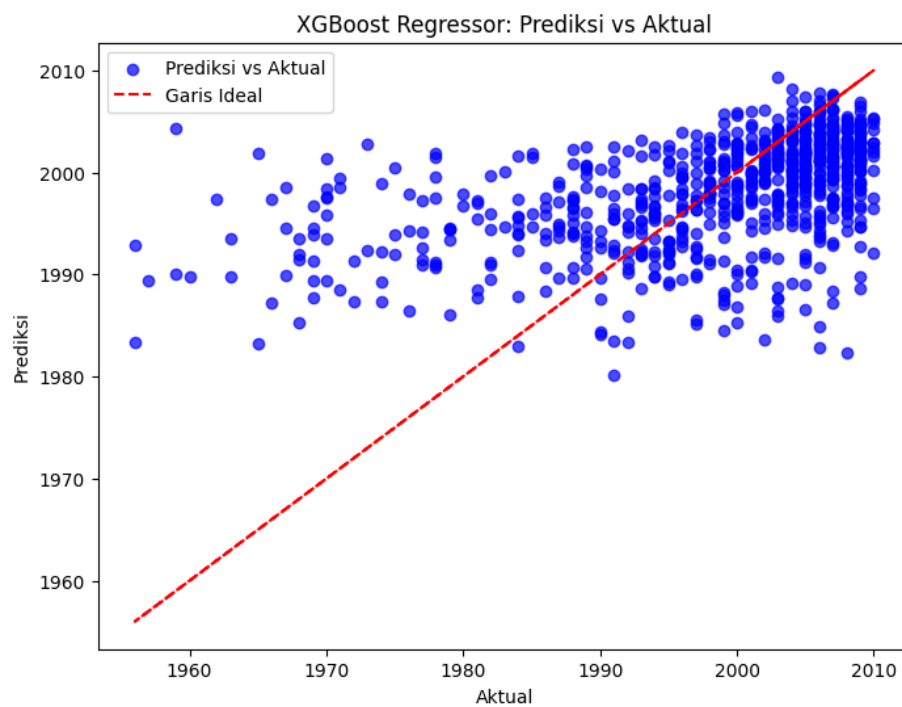
k-NN Regressor: Prediksi vs Aktual

Laporan Evaluasi:

Best Parameters: {'regressor__colsample_bytree': 0.6, 'regressor__learning_rate': 0.1, 'regressor__max_depth': 3, 'regressor__n_estimators': 100, 'regressor__subsample': 1.0}

MSE: 91.07851157441735

R2: 0.2318658056199524

XGBoost Regressor: Prediksi vs Aktual

Lima baris pertama dataset:

    Clump_thickness Uniformity_of_cell_size Uniformity_of_cell_shape \

	Clump_thickness	Uniformity_of_cell_size	Uniformity_of_cell_shape
0	5	1	1
1	5	4	4
2	3	1	1
3	6	8	8
4	4	1	1

	Marginal_adhesion	Single_epithelial_cell_size	Bare_nuclei \
0	1	2	1.0
1	5	7	10.0
2	1	2	2.0
3	1	3	4.0
4	3	2	1.0

	Bland_chromatin	Normal_nucleoli	Mitoses	Class
0	3	1	1	2
1	3	2	1	2
2	3	1	1	2
3	3	7	1	2
4	3	1	1	2

Informasi dataset:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 699 entries, 0 to 698

Data columns (total 10 columns):

#	Column	Non-Null Count	Dtype
---	------	--------------	-----
0	Clump_thickness	699 non-null	int64
1	Uniformity_of_cell_size	699 non-null	int64
2	Uniformity_of_cell_shape	699 non-null	int64
3	Marginal_adhesion	699 non-null	int64
4	Single_epithelial_cell_size	699 non-null	int64
5	Bare_nuclei	683 non-null	float64

```
 6   Bland_chromatin          699 non-null    int64
 7   Normal_nucleoli          699 non-null    int64
 8   Mitoses                  699 non-null    int64
 9   Class                    699 non-null    int64
dtypes: float64(1), int64(9)
memory usage: 54.7 KB
None
```

Statistik deskriptif dataset:

	Clump_thickness	Uniformity_of_cell_size	Uniformity_of_cell_shape \
count	699.000000	699.000000	699.000000
mean	4.417740	3.134478	3.207439
std	2.815741	3.051459	2.971913
min	1.000000	1.000000	1.000000
25%	2.000000	1.000000	1.000000
50%	4.000000	1.000000	1.000000
75%	6.000000	5.000000	5.000000
max	10.000000	10.000000	10.000000

	Marginal_adhesion	Single_epithelial_cell_size	Bare_nuclei \
count	699.000000	699.000000	683.000000
mean	2.806867	3.216023	3.544656
std	2.855379	2.214300	3.643857
min	1.000000	1.000000	1.000000
25%	1.000000	2.000000	1.000000
50%	1.000000	2.000000	1.000000
75%	4.000000	4.000000	6.000000
max	10.000000	10.000000	10.000000

	Bland_chromatin	Normal_nucleoli	Mitoses	Class
count	699.000000	699.000000	699.000000	699.000000
mean	3.437768	2.866953	1.589413	2.689557

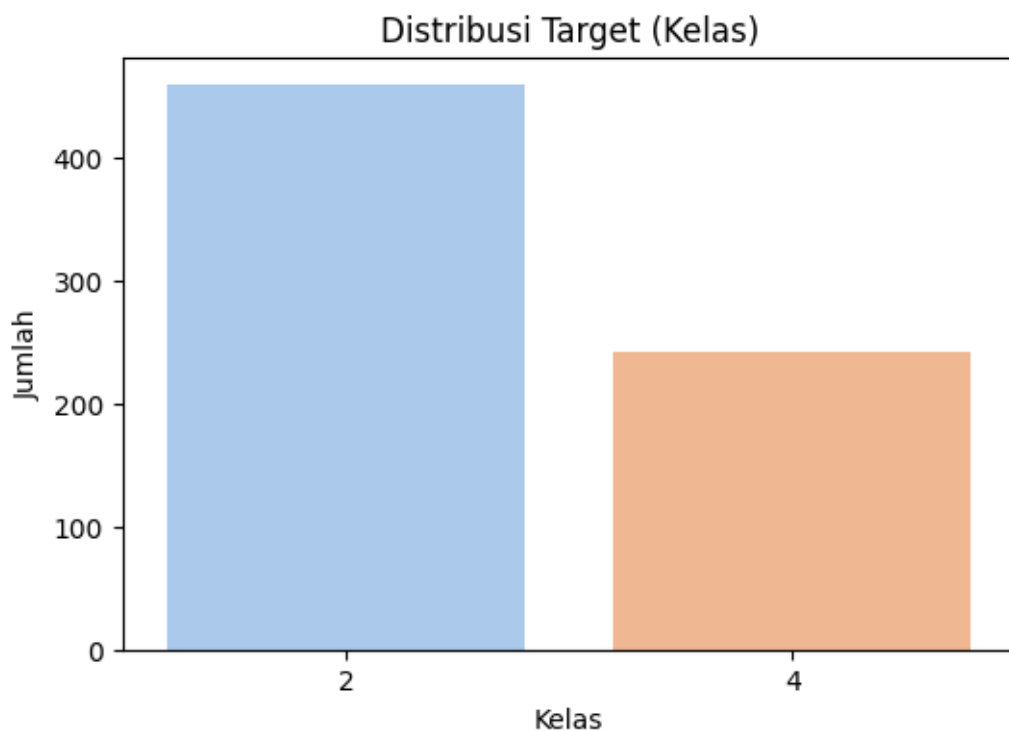| std	2.438364	3.053634	1.715078	0.951273
min	1.000000	1.000000	1.000000	2.000000
25%	2.000000	1.000000	1.000000	2.000000
50%	3.000000	1.000000	1.000000	2.000000
75%	5.000000	4.000000	1.000000	4.000000
max	10.000000	10.000000	10.000000	4.000000

Distribusi target:

Class

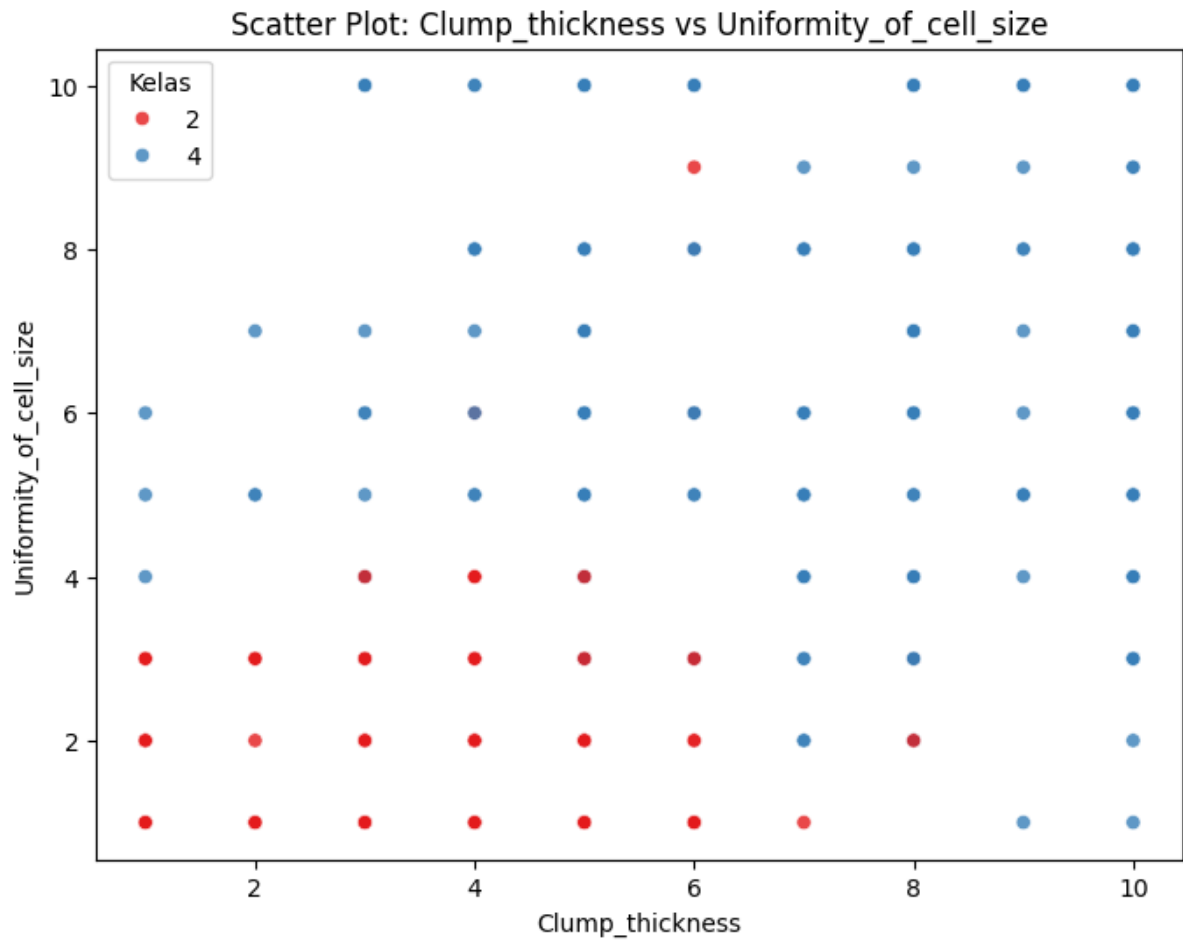2   458

4   241

Name: count, dtype: int64

<ipython-input-2-1c45595c9e5d>:37: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.countplot(x=target_column, data=data, palette="pastel")
```


Distribusi Target (Kelas)

Heatmap Korelasi Antar Variabel

Distribusi Fitur Berdasarkan Target

Scatter Plot: Clump_thickness vs Uniformity_of_cell_size
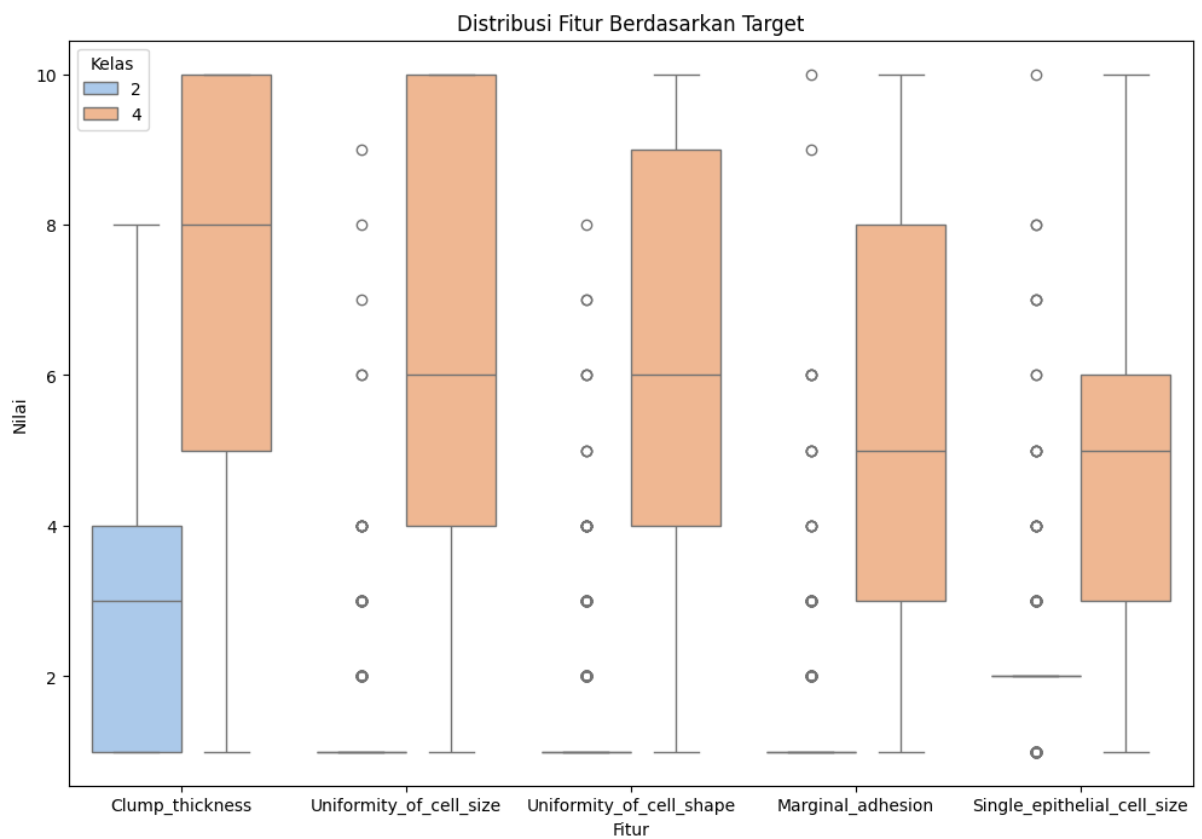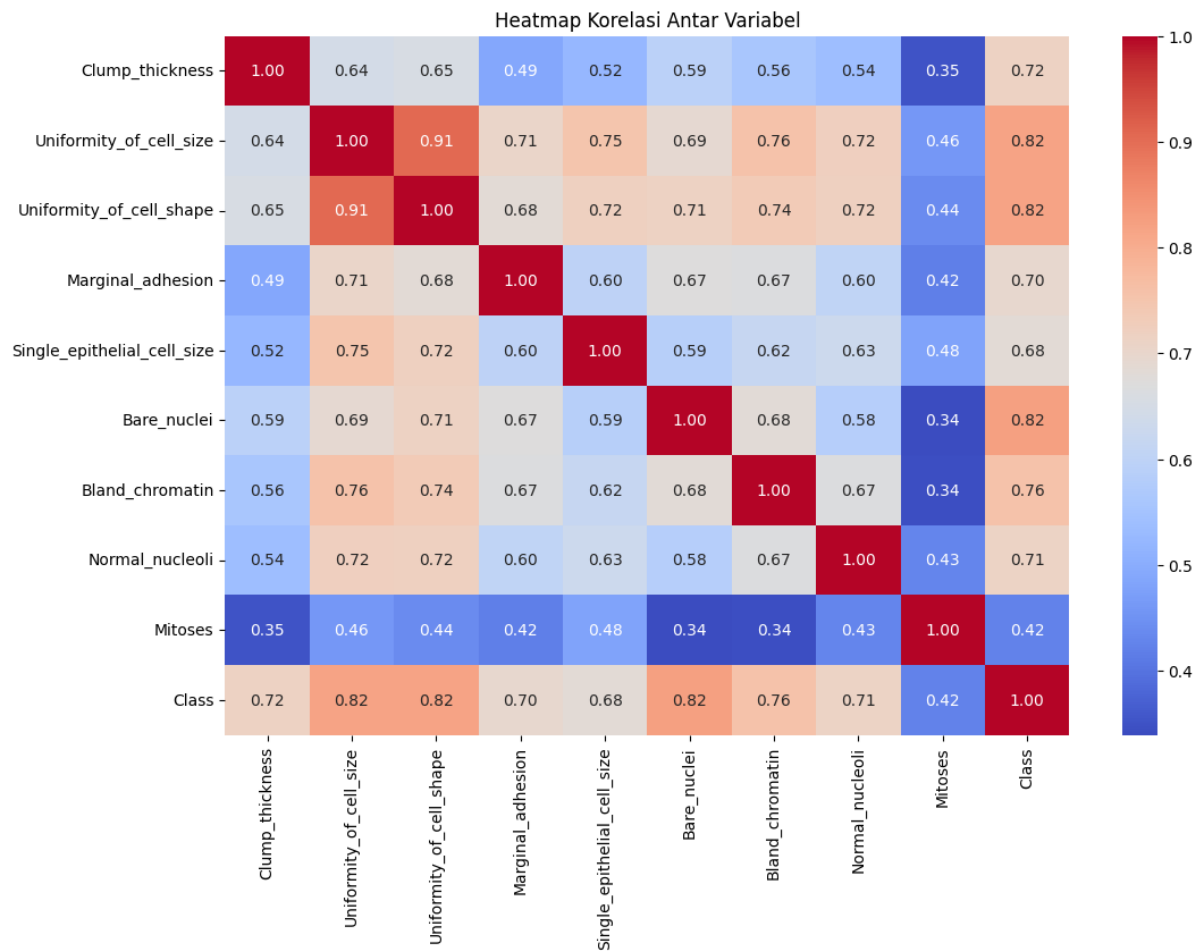
Laporan Evaluasi:

	precision	recall	f1-score	support
2	0.96	0.97	0.96	92
4	0.94	0.92	0.93	48
accuracy			0.95	140
macro avg	0.95	0.94	0.94	140
weighted avg	0.95	0.95	0.95	140

Parameter Terbaik:

{'classifier__C': 1, 'classifier__penalty': 'l1', 'classifier__solver': 'liblinear'}
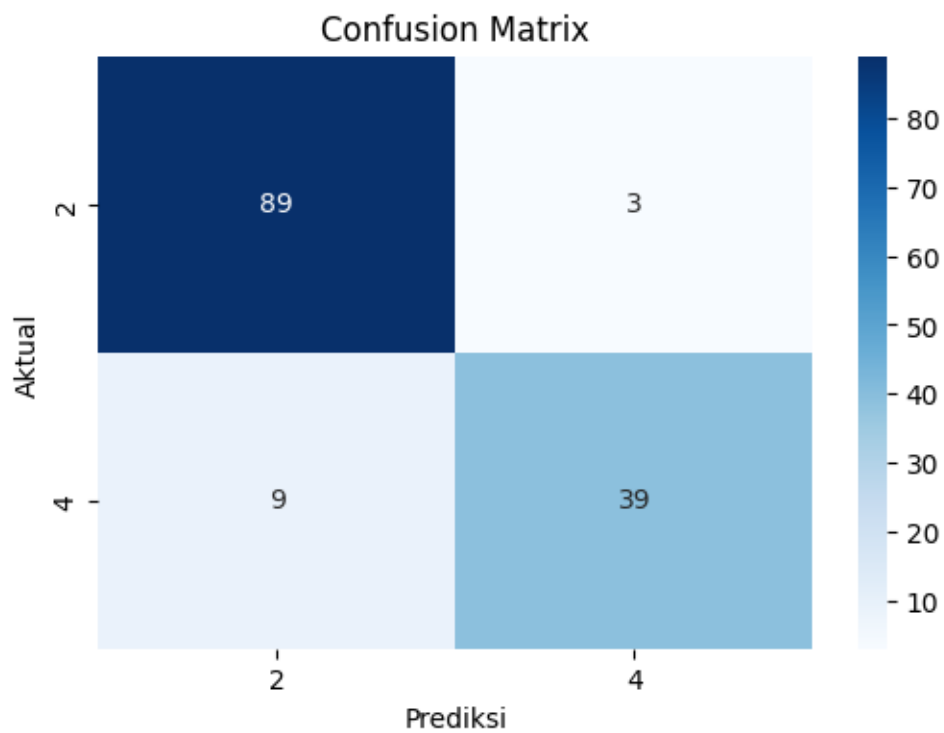
Confusion Matrix
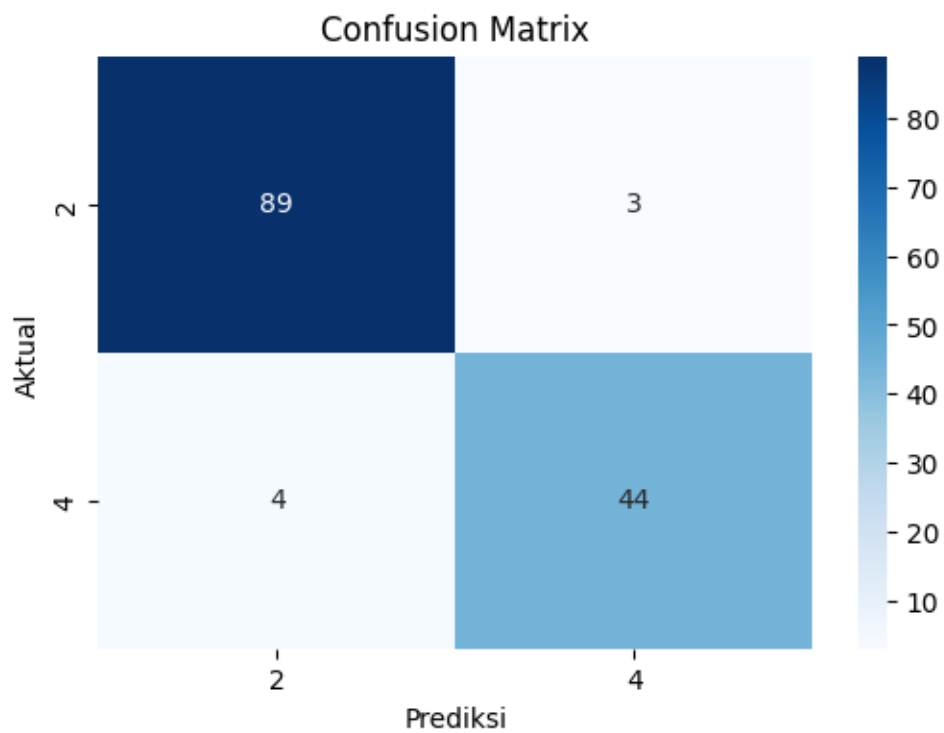
Laporan Evaluasi:

	precision	recall	f1-score	support
2	0.91	0.97	0.94	92
4	0.93	0.81	0.87	48
accuracy			0.91	140
macro avg	0.92	0.89	0.90	140
weighted avg	0.92	0.91	0.91	140

Parameter Terbaik:

{'classifier__criterion': 'entropy', 'classifier__max_depth': 5, 'classifier__max_features': None, 'classifier__min_samples_leaf': 1, 'classifier__min_samples_split': 5}
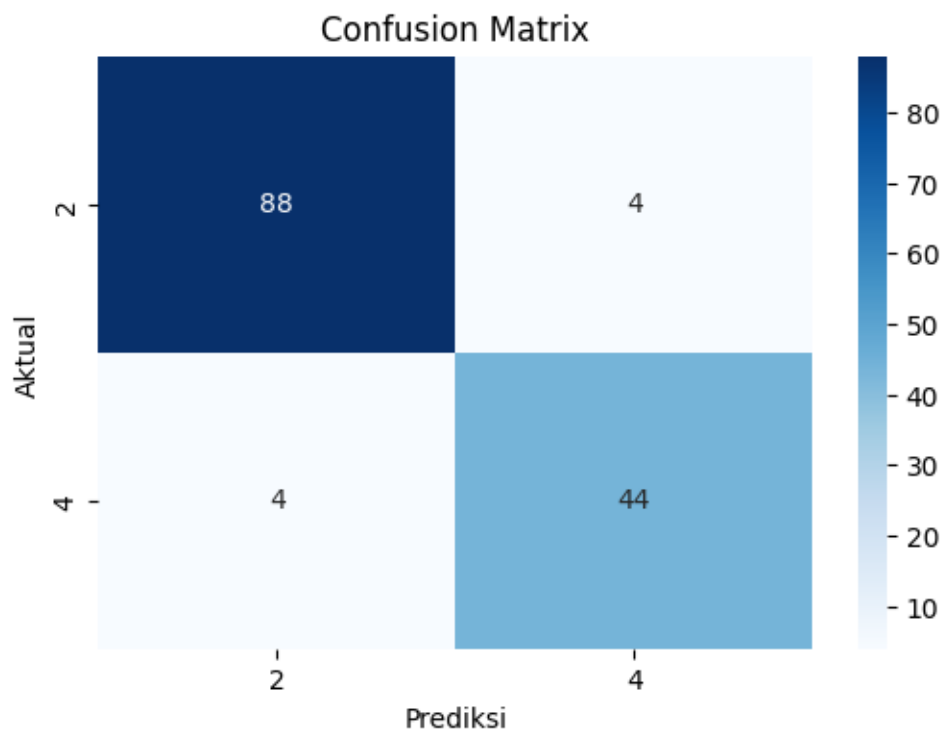
Confusion Matrix

Laporan Evaluasi:

	precision	recall	f1-score	support
2	0.96	0.96	0.96	92
4	0.92	0.92	0.92	48
accuracy			0.94	140
macro avg	0.94	0.94	0.94	140
weighted avg	0.94	0.94	0.94	140

Parameter Terbaik:

{'classifier__metric': 'manhattan', 'classifier__n_neighbors': 3, 'classifier__p': 1, 'classifier__weights': 'distance'}

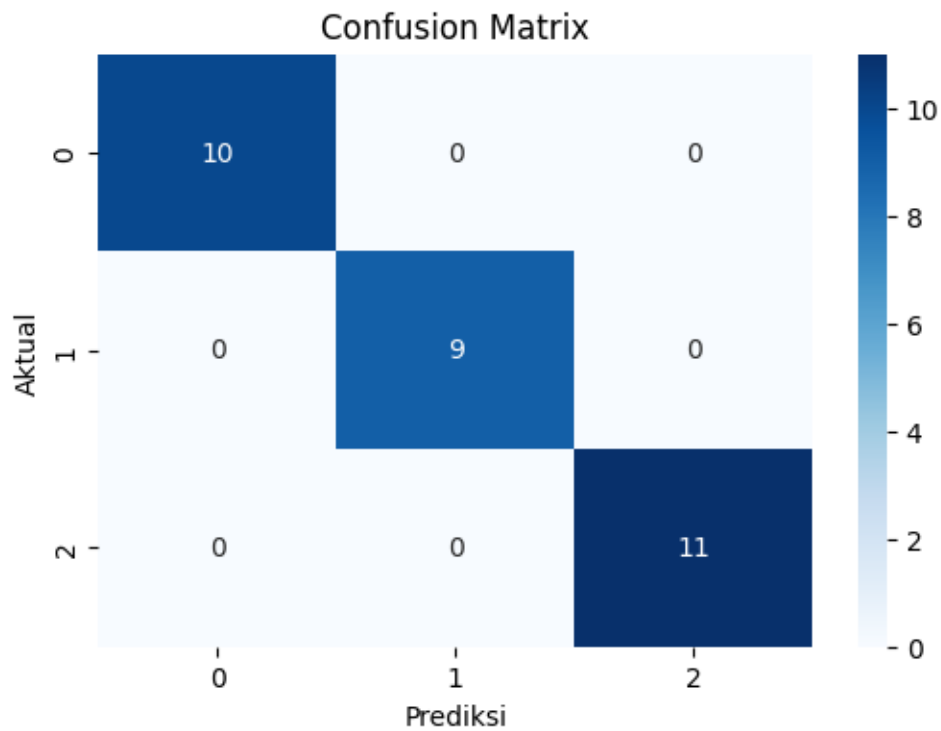Confusion Matrix

Laporan Evaluasi:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	10
1	1.00	1.00	1.00	9
2	1.00	1.00	1.00	11
accuracy			1.00	30
macro avg	1.00	1.00	1.00	30
weighted avg	1.00	1.00	1.00	30

Parameter Terbaik:

{'classifier__colsample_bytree': 1.0, 'classifier__learning_rate': 0.01, 'classifier__max_depth': 3, 'classifier__n_estimators': 200, 'classifier__subsample': 1.0}

Laporan Evaluasi:

	precision	recall	f1-score	support
2	0.97	0.98	0.98	143
4	0.95	0.94	0.95	67
accuracy			0.97	210
macro avg	0.96	0.96	0.96	210
weighted avg	0.97	0.97	0.97	210

Parameter Terbaik:

{'xgb__learning_rate': 0.1, 'xgb__max_depth': 5, 'xgb__n_estimators': 100}

Confusion Matrix