

## ✓ UTS Machine Learning - Tugas Regresi

Nama: M Faishal Abdurrahman

NIM: 110323015

Struktur Awal

```
# UTS Machine Learning - Tugas Regresi
# Nama: M Faishal Abdurrahman
# NIM: 110323015

# Import Library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
import warnings
warnings.filterwarnings('ignore')

# Mount Google Drive
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

Load dan Eksplorasi Data

```
# Load Dataset
path_regresi = '/content/drive/My Drive/UTS/RegresiUTSTelkom.csv'

# Untuk dataset besar, gunakan sampel terlebih dahulu
df_regresi = pd.read_csv(path_regresi)
# Alternatif jika dataset terlalu besar:
# df_regresi = pd.read_csv(path_regresi, nrows=50000) # Baca sebagian data saja

# Eksplorasi Data
print("Informasi Dataset:")
print(df_regresi.info())

print("\nStatistik Deskriptif:")
print(df_regresi.describe())

print("\nJumlah data missing:")
print(df_regresi.isnull().sum())

# Tampilkan beberapa baris pertama
print("\nSampel Data:")
df_regresi.head()
```

↳ **Informasi Dataset:**  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 515344 entries, 0 to 515343  
Data columns (total 91 columns):  
 # Column Non-Null Count Dtype  
--- ---  
 0 2001 515344 non-null int64  
 1 49.94357 515344 non-null float64  
 2 21.47114 515344 non-null float64  
 3 73.0775 515344 non-null float64  
 4 8.74861 515344 non-null float64  
 5 -17.40628 515344 non-null float64  
 6 -13.09905 515344 non-null float64  
 7 -25.01202 515344 non-null float64  
 8 -12.23257 515344 non-null float64  
 9 7.83089 515344 non-null float64  
 10 -2.46783 515344 non-null float64  
 11 3.32136 515344 non-null float64  
 12 -2.31521 515344 non-null float64  
 13 10.20556 515344 non-null float64  
 14 611.10913 515344 non-null float64  
 15 951.0896 515344 non-null float64  
 16 698.11428 515344 non-null float64  
 17 408.98485 515344 non-null float64  
 18 383.70912 515344 non-null float64  
 19 326.51512 515344 non-null float64  
 20 238.11327 515344 non-null float64  
 21 251.42414 515344 non-null float64  
 22 187.17351 515344 non-null float64  
 23 100.42652 515344 non-null float64  
 24 179.19498 515344 non-null float64  
 25 -8.41558 515344 non-null float64  
 26 -317.87038 515344 non-null float64  
 27 95.86266 515344 non-null float64  
 28 48.10259 515344 non-null float64  
 29 -95.66303 515344 non-null float64  
 30 -18.06215 515344 non-null float64  
 31 1.96984 515344 non-null float64  
 32 34.42438 515344 non-null float64  
 33 11.7267 515344 non-null float64  
 34 1.3679 515344 non-null float64  
 35 7.79444 515344 non-null float64  
 36 -0.36994 515344 non-null float64  
 37 -133.67852 515344 non-null float64  
 38 -83.26165 515344 non-null float64  
 39 -37.29765 515344 non-null float64  
 40 73.04667 515344 non-null float64  
 41 -37.36684 515344 non-null float64  
 42 -3.13853 515344 non-null float64  
 43 -24.21531 515344 non-null float64  
 44 -13.23866 515344 non-null float64  
 45 15.93809 515344 non-null float64  
 46 -18.60478 515344 non-null float64  
 47 82.15479 515344 non-null float64  
 48 240.5798 515344 non-null float64  
 49 -10.29407 515344 non-null float64  
 50 31.58431 515344 non-null float64  
 51 -25.38187 515344 non-null float64  
 52 -3.90772 515344 non-null float64  
 53 13.29258 515344 non-null float64  
 54 41.5506 515344 non-null float64  
 55 -7.26272 515344 non-null float64  
 56 -21.00863 515344 non-null float64  
 57 105.50848 515344 non-null float64  
 58 64.29856 515344 non-null float64  
 59 26.08481 515344 non-null float64  
 60 -44.5911 515344 non-null float64  
 61 -8.30657 515344 non-null float64  
 62 7.93706 515344 non-null float64  
 63 -10.7366 515344 non-null float64  
 64 -95.44766 515344 non-null float64  
 65 -82.03307 515344 non-null float64  
 66 -35.59194 515344 non-null float64  
 67 4.69525 515344 non-null float64  
 68 70.95626 515344 non-null float64  
 69 28.09139 515344 non-null float64  
 70 6.02015 515344 non-null float64  
 71 -37.13767 515344 non-null float64  
 72 -41.1245 515344 non-null float64  
 73 -8.40816 515344 non-null float64  
 74 7.19877 515344 non-null float64  
 75 -8.60176 515344 non-null float64  
 76 -5.90857 515344 non-null float64  
 77 -12.32437 515344 non-null float64  
 78 14.68734 515344 non-null float64  
 79 -54.32125 515344 non-null float64  
 80 40.14786 515344 non-null float64  
 81 13.0162 515344 non-null float64  
 82 -54.40548 515344 non-null float64  
 83 58.99367 515344 non-null float64

```

84 15.37344 515344 non-null float64
85 1.11144 515344 non-null float64
86 -23.08793 515344 non-null float64
87 68.40795 515344 non-null float64
88 -1.82223 515344 non-null float64
89 -27.46348 515344 non-null float64
90 2.26327 515344 non-null float64
dtypes: float64(90), int64(1)
memory usage: 357.8 MB
None

```

Statistik Deskriptif:

	2001	49.94357	21.47114	73.0775	\
count	515344.000000	515344.000000	515344.000000	515344.000000	
mean	1998.397077	43.387113	1.289515	8.658222	
std	10.931056	6.067557	51.580393	35.268505	
min	1922.000000	1.749000	-337.092500	-301.005060	
25%	1994.000000	39.954667	-26.059848	-11.462775	
50%	2002.000000	44.258490	8.417725	10.476235	
75%	2006.000000	47.833875	36.124030	29.764685	
max	2011.000000	61.970140	384.065730	322.851430	
	8.74861	-17.40628	-13.09905	-25.01202	\
count	515344.000000	515344.000000	515344.000000	515344.000000	
mean	1.164110	-6.553580	-9.521968	-2.391046	
std	16.322802	22.860803	12.857763	14.571853	
min	-154.183580	-181.953370	-81.794290	-188.214000	
25%	-8.487507	-20.666455	-18.441005	-10.780360	
50%	-0.652855	-6.007770	-11.188355	-2.046625	
75%	8.787548	7.741877	-2.388945	6.508587	
max	335.771820	262.068870	166.236890	172.402680	
	-12.23257	7.83089	...	13.0162	-54.40548 \
count	515344.000000	515344.000000	...	515344.000000	515344.000000
mean	-1.793215	3.727868	...	15.755411	-73.461537
std	7.963822	10.582869	...	32.099666	175.619058
min	-72.503850	-126.479040	...	-437.722030	-4402.376440
25%	-6.468390	-2.293670	...	-1.812658	-139.555737
50%	-1.736415	3.822305	...	9.171850	-53.089115
75%	2.913455	9.961865	...	26.274487	13.478793
max	126.741270	146.297950	...	840.973380	4469.454870
	58.99367	15.37344	1.11144	-23.08793	\
count	515344.000000	515344.000000	515344.000000	515344.000000	
mean	41.542388	37.934163	0.315750	17.669292	
std	122.228915	95.050718	16.161780	114.428002	
min	-1810.689190	-3098.350310	-341.789120	-3168.924570	
25%	-20.987115	-4.669655	-6.781598	-31.580617	
50%	28.790580	33.623815	0.820830	15.598520	
75%	89.661785	77.785810	8.471000	67.795110	
max	3210.701700	1734.079690	260.544900	3662.065650	
	68.40795	-1.82223	-27.46348	2.26327	
count	515344.000000	515344.000000	515344.000000	515344.000000	
mean	-26.315520	4.458653	20.035229	1.329104	
std	173.977455	13.346567	185.558415	22.088598	
min	-4319.992320	-236.039260	-7458.378150	-381.424430	
25%	-101.530305	-2.566137	-59.509453	-8.820248	
50%	-21.204225	3.117645	7.759910	0.053015	
75%	52.389322	9.967742	86.351645	9.679540	
max	2833.608950	463.419500	7393.398440	677.899630	

[8 rows x 91 columns]

Jumlah data missing:

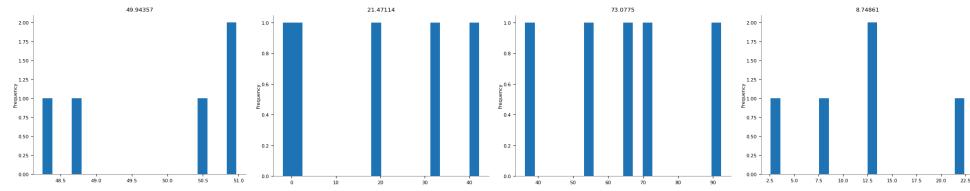
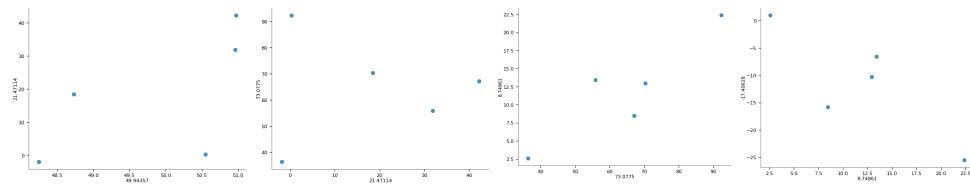
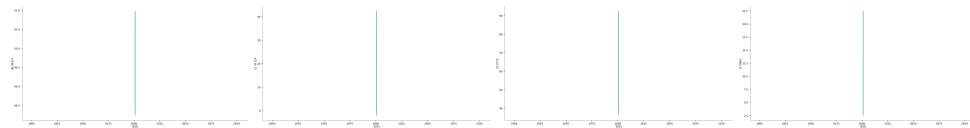
2001	0
49.94357	0
21.47114	0
73.0775	0
8.74861	0
	..
-23.08793	0
68.40795	0
-1.82223	0
-27.46348	0
2.26327	0

Length: 91, dtype: int64

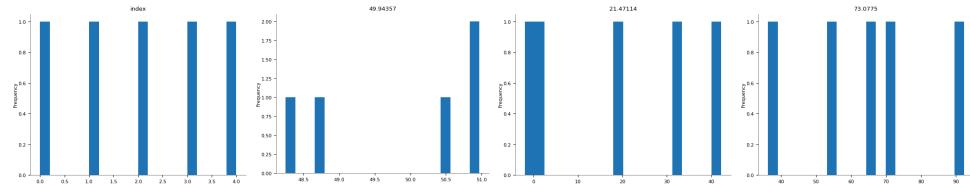
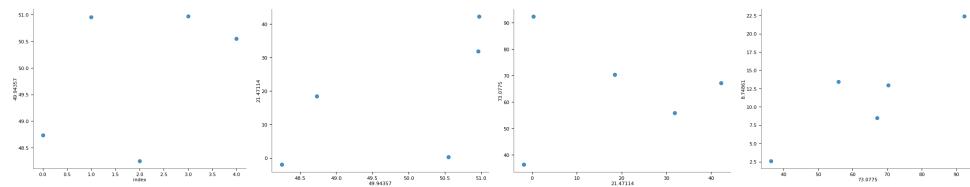
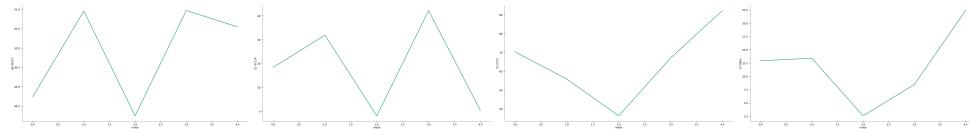
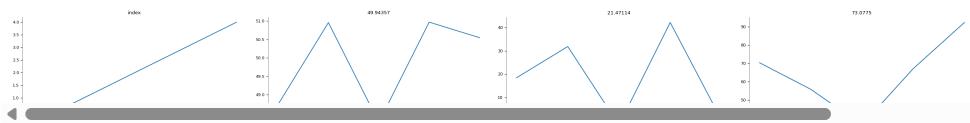
Sampel Data:

	2001	49.94357	21.47114	73.0775	8.74861	-17.40628	-13.09905	-25.01202	-12.23257	7.83089	...	13.0162	-54.40548	58.
0	2001	48.73215	18.42930	70.32679	12.94636	-10.32437	-24.83777	8.76630	-0.92019	18.76548	...	5.66812	-19.68073	33.
1	2001	50.95714	31.85602	55.81851	13.41693	-6.57898	-18.54940	-3.27872	-2.35035	16.07017	...	3.03800	26.05866	-50.
2	2001	48.24750	-1.89837	36.29772	2.58776	0.97170	-26.21683	5.05097	-10.34124	3.55005	...	34.57337	-171.70734	-16.
3	2001	50.97020	42.20998	67.09964	8.46791	-15.85279	-16.81409	-12.48207	-9.37636	12.63699	...	9.92661	-55.95724	64.
4	2001	50.54767	0.31568	92.35066	22.38696	-25.51870	-19.04928	20.67345	-5.19943	3.63566	...	6.59753	-50.69577	26.

5 rows x 91 columns

**Distributions****2-d distributions****Time series**

Warning: Total number of columns (91) exceeds max\_columns (20) limiting to first (20) columns.

**Distributions****2-d distributions****Time series****Values**

## Visualisasi Data

```
import matplotlib.pyplot as plt
import seaborn as sns

target = '2001'
features = [c for c in df_regresi.columns if c not in ['index', target]]
df_plot = df_regresi[features + [target]].astype(float)

# 1) Distribusi target
plt.figure(figsize=(12,6))
sns.histplot(df_plot[target], kde=True)
plt.title(f'Distribusi {target}')
plt.xlabel(target)
plt.ylabel('Frekuensi')
plt.show()

# 2) Heatmap korelasi
plt.figure(figsize=(10,8))
corr = df_plot.corr()
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Matriks Korelasi')
plt.show()

# 3) Scatter plot tiap fitur vs target
plt.figure(figsize=(16, 4*len(features)))
for i, feat in enumerate(features):
    plt.subplot(len(features), 1, i+1)
    sns.scatterplot(x=df_plot[feat], y=df_plot[target])
    plt.xlabel(feat)
    plt.ylabel(target)
    plt.title(f'{feat} vs {target}')
plt.tight_layout()
plt.show()
```

