Predict Electricity From Fuels

Final Project - Renewenergy

Swipe

Profil Team

Noviana Rosianti
Lead
Universitas Pancasila

Naomi Dwi Anggraini
Data Analyst
Universitas Pembangunan Nasional Veteran Jatim

Sadam Mochamad Fahrurozi
Data Analyst
Universitas Islam Bandung

Mohamad Irfan
Data Visualization
Universitas Tadulako

Wulandari Tri Rahayu
Data Visualization
Universitas Negeri Padang

Project Overview

01 Business Understanding

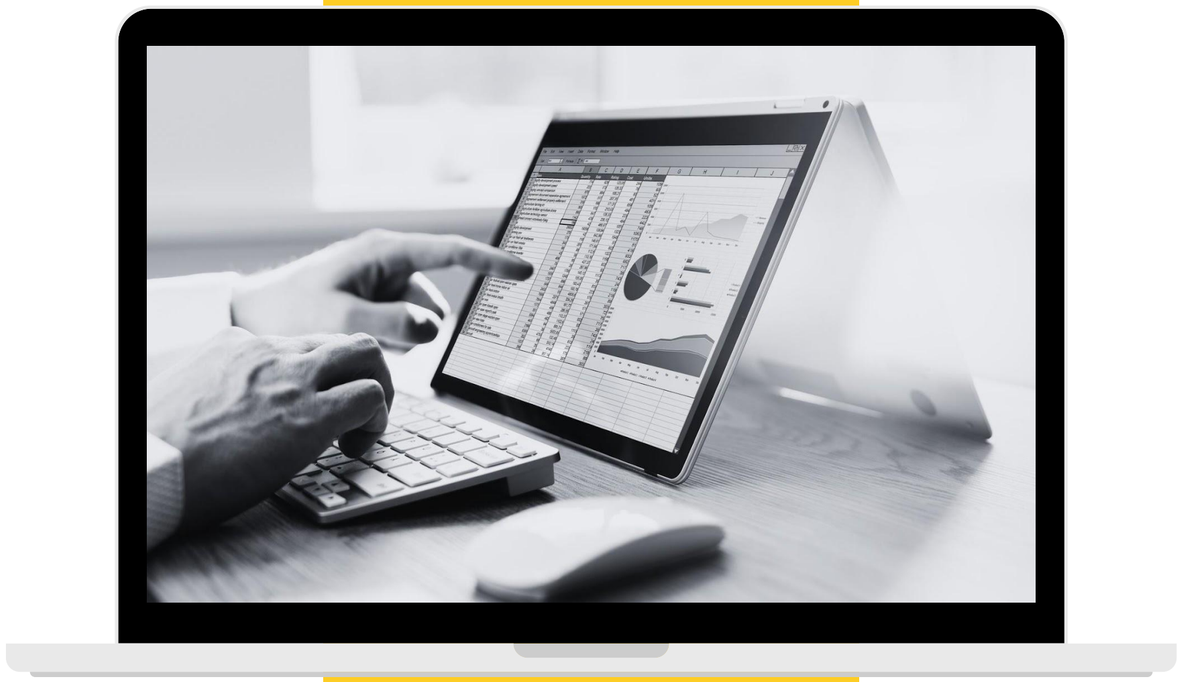
02 Data Understanding

03 Data Preparation

04 Modelling

05 Evaluasi

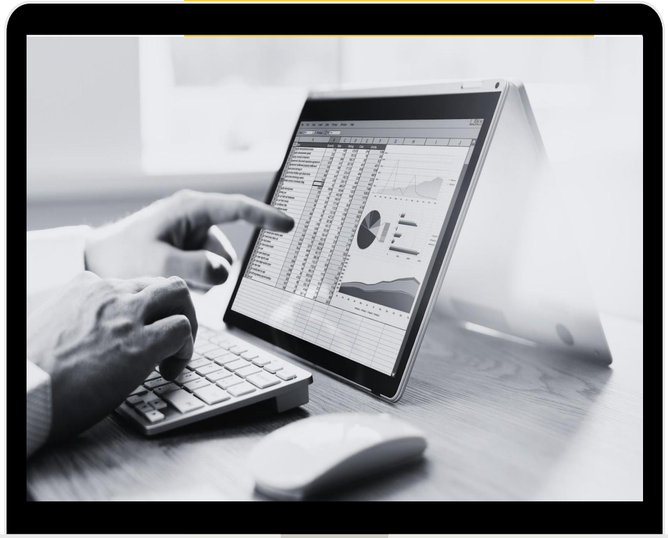
06 Deployment



Business Understanding

Business Understanding

1. Permintaan energi global dari sumber non-terbarukan meningkat.
2. Analisis data produksi listrik dari bahan bakar fosil selama dua dekade terakhir (2000-2020) dan informasi mengenai konsekuensi lingkungan adalah kunci dalam merumuskan strategi yang lebih berkelanjutan.
3. Sumber-sumber seperti Our World in Data memberikan wawasan penting tentang keberlanjutan energi dan lingkungan, membantu bisnis mengatasi tantangan lingkungan akibat bahan bakar fosil yang tidak berkelanjutan.

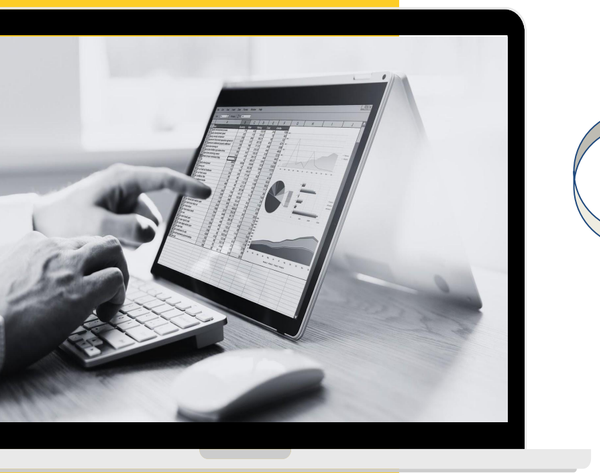


Website

www.greatedu.co.id

Tujuan & Manfaat

- Memprediksi produksi listrik yang berasal dari bahan bakar fosil



Dengan memahami prediksi listrik yang berasal dari bahan bakar fosil maka dapat di identifikasi tren konsumsi listrik yang berasal dari bahan bakar fosil sehingga manfaat yang dapat diperoleh:

- Membantu pengambilan keputusan oleh pemangku kepentingan untuk merencanakan kebijakan, investasi, atau inisiatif yang lebih baik terkait energi dan lingkungan.
- Mengurangi dampak lingkungan dari konsumsi energi berbasis bahan bakar fosil.

Website

www.greatedu.co.id

Teknik dan Tools

Metode yang digunakan untuk membuat prediksi dari tren konsumsi energi global adalah metode regresi (regression).

Pandas, NumPy ⇨ manipulasi dan analisis data

matplotlib, seaborn ⇨ visualisasi

scikit learn ⇨ modelling

Algoritma yang digunakan untuk pemodelan adalah Random Forest Regressor, Linear Regression, dan Gradient Boosting Regressor. Evaluasi kinerja model dilakukan dengan menggunakan RMSE serta pembuatan dashboard dilakukan dengan menggunakan Google Looker Studio.

Website

www.greatedu.co.id

Data Understanding

Swipe

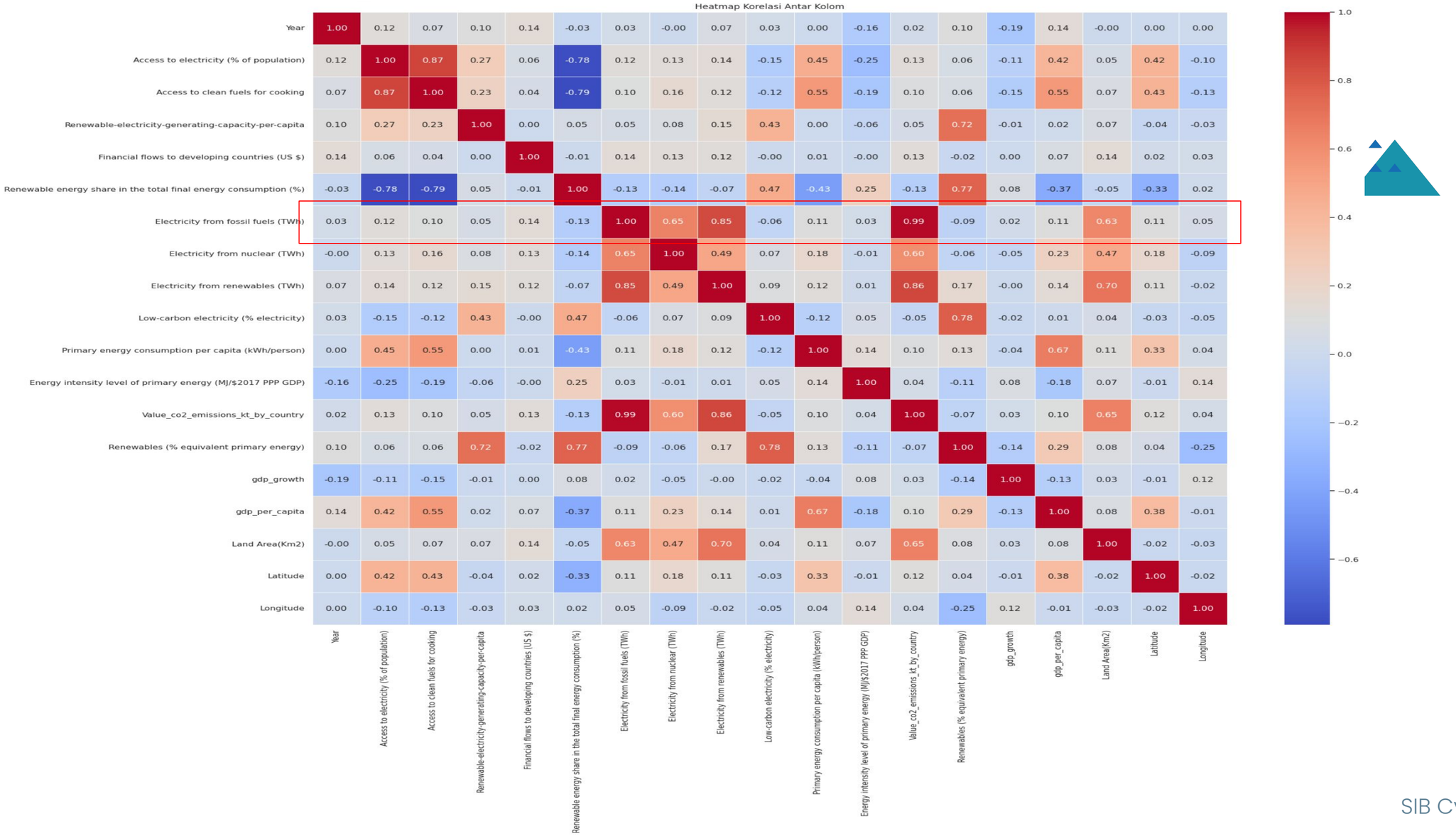
Menggunakan dataset global_data_on_sustainable_energy merupakan kumpulan data bersumber dari situs https://ourworldindata.org/sdgs/affordable-clean-energy. Data berisi indikator energi berkelanjutan dan faktor-faktor lainnya seperti akses listrik, energi terbarukan, emisi karbon, dan intensitas energi di beberapa negara dari tahun 2000 hingga 2020. Data terdiri dari 3649 baris dan 21 kolom.

Variabel	Tipe Data	Keterangan
Entity	Object	Nama negara atau wilayah untuk yang data tersebut dilaporkan.
Year	Integer	Tahun untuk yang data tersebut dilaporkan, berkisar dari 2000 hingga 2020.
Access to electricity (% of population)	Float	Persentase dari populasi yang memiliki akses listrik.
Access to clean fuels for cooking (% of population)	Float	Persentase dari populasi yang bergantung pada bahan bakar bersih utama.

Renewable-electricity-generating-capacity-per-capita	Float	Kapasitas energi terbarukan yang diinstal per orang.
Financial flows to developing countries (US $)	Float	Bantuan dan bantuan dari negara maju untuk proyek energi bersih.
Renewable energi share in total final energy consumption (%)	Float	Persentase energi terbarukan dalam konsumsi energi akhir.
Electricity from fossil fuels (TWh)	Float	Listrik yang dihasilkan dari bahan bakar fosil (batu bara, minyak, gas) dalam terawatt-jam.
Electricity from nuclear (TWh)	Float	Listrik yang dihasilkan dari tenaga nuklir dalam terawatt-jam.
Electricity from renewables (TWh)	Float	Listrik yang dihasilkan dari sumber terbarukan (hidro, surya, angin, dll.) dalam terawatt-jam.
Low-carbon electricity (% electricity)	Float	Persentase listrik dari sumber rendah karbon (nuklir dan terbarukan).
Primary energy consumption per capita (kWh/person)	Float	Konsumsi energi per orang dalam kilowatt-jam.

Energy intensity level of primary energy (MJ/$2011 PPP GDP)	Float	Penggunaan energi per unit GDP pada paritas daya beli.
Value_co2_emissions (metric tons per capita)	Float	Emisi karbon dioksida per orang dalam ton metrik.
Renewables (% equivalent primary energy)	Float	Energi primer setara yang berasal dari sumber terbarukan.
GDP growth (annual %)	Float	Tingkat pertumbuhan PDB tahunan berdasarkan mata uang lokal konstan.
GDP per capita	Float	Produk domestik bruto per orang.
Density (P/Km2)	Object	Kepadatan penduduk dalam orang per kilometer persegi.
Land Area (Km2)	Float	Luas total tanah dalam kilometer persegi.
Latitude	Float	Lintang pusat negara dalam derajat desimal.
Longitude	Float	Bujur pusat negara dalam derajat desimal.

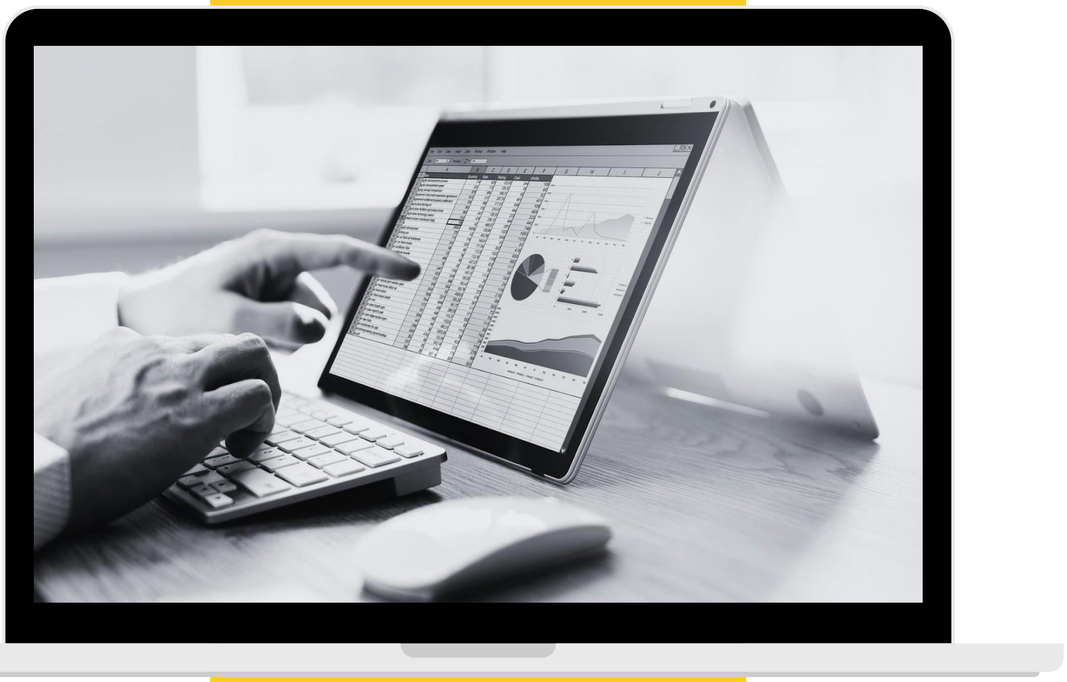
Heatmap Correlation



Korelasi antar atribut

	Year	Access to electricity (% of population)	Access to clean fuels for cooking	Renewable-electricity-generating-capacity-per-capita	Financial flows to developing countries (US $)	Renewable energy share in the total final energy consumption (%)	Electricity from fossil fuels (TWh)	Electricity from nuclear (TWh)	Electricity from renewables (TWh)	Low-carbon electricity (% electricity)	Primary energy consumption per capita (kWh/person)	Energy intensity level of primary energy (MJ/$2017 PPP GDP)	Value_co2_emissions_kt_by_country	Renewables (% equivalent primary energy)	gdp_growth	gdp_per_capita	Land Area(Km2)	Latitude	Longitude
Year	1.000000	0.124890	0.073292	0.104365	0.141820	-0.029384	0.032615	-0.004820	0.073935	0.030126	0.004196	-0.161605	0.024759	0.104767	-0.192566	0.135399	-0.001629	0.003507	0.000944
Access to electricity (% of population)	0.124890	1.000000	0.866554	0.270423	0.060478	-0.784514	0.123999	0.128713	0.138057	-0.151062	0.454870	-0.252949	0.126988	0.058195	-0.112071	0.418459	0.054964	0.420548	-0.102857
Access to clean fuels for cooking	0.073292	0.866554	1.000000	0.227962	0.042204	-0.791740	0.100752	0.163473	0.124486	-0.120110	0.551347	-0.194884	0.095629	0.063131	-0.145301	0.545165	0.074630	0.432700	-0.131847
Renewable-electricity-generating-capacity-per-capita	0.104365	0.270423	0.227962	1.000000	0.003677	0.054821	0.051483	0.080168	0.147768	0.427264	0.000290	-0.056445	0.046755	0.718511	-0.013462	0.017266	0.070355	-0.041691	-0.025587
Financial flows to developing countries (US $)	0.141820	0.060478	0.042204	0.003677	1.000000	-0.005519	0.139238	0.127840	0.122362	-0.002716	0.009860	-0.004231	0.126609	-0.023016	0.003324	0.067525	0.141528	0.021782	0.029141
Renewable energy share in the total final energy consumption (%)	-0.029384	-0.784514	-0.791740	0.054821	-0.005519	1.000000	-0.134888	-0.136525	-0.068779	0.467862	-0.429362	0.247050	-0.132715	0.768166	0.075236	-0.370648	-0.045395	-0.330140	0.018284
Electricity from fossil fuels (TWh)	0.032615	0.123999	0.100752	0.051483	0.139238	-0.134888	1.000000	0.646716	0.845067	-0.063773	0.105529	0.030998	0.989741	-0.088586	0.020956	0.107793	0.627682	0.110994	0.045413
Electricity from nuclear (TWh)	-0.004820	0.128713	0.163473	0.080168	0.127840	-0.136525	0.646716	1.000000	0.492011	0.071715	0.180833	-0.007496	0.598495	-0.057667	-0.048455	0.231349	0.473280	0.177961	-0.087111
Electricity from renewables (TWh)	0.073935	0.138057	0.124486	0.147768	0.122362	-0.068779	0.845067	0.492011	1.000000	0.091451	0.119760	0.009425	0.856270	0.166374	-0.002531	0.139674	0.704884	0.112833	-0.016380
Low-carbon electricity (% electricity)	0.030126	-0.151062	-0.120110	0.427264	-0.002716	0.467862	-0.063773	0.071715	0.091451	1.000000	-0.119386	0.047489	-0.054515	0.782388	-0.015196	0.007945	0.044866	-0.029134	-0.045336
Primary energy consumption per capita (kWh/person)	0.004196	0.454870	0.551347	0.000290	0.009860	-0.429362	0.105529	0.180833	0.119760	-0.119386	1.000000	0.137876	0.095683	0.130623	-0.035434	0.667172	0.114661	0.329893	0.039724
Energy intensity level of primary energy (MJ/$2017 PPP GDP)	-0.161605	-0.252949	-0.194884	-0.056445	-0.004231	0.247050	0.030998	-0.007496	0.009425	0.047489	0.137876	1.000000	0.039128	-0.106122	0.079293	-0.184681	0.071686	-0.014179	0.135104
Value_co2_emissions_kt_by_country	0.024759	0.126988	0.095629	0.046755	0.126609	-0.132715	0.989741	0.598495	0.856270	-0.054515	0.095683	0.039128	1.000000	-0.073735	0.031523	0.098209	0.653732	0.115571	0.036286
Renewables (% equivalent primary energy)	0.104767	0.058195	0.063131	0.718511	-0.023016	0.768166	-0.088586	-0.057667	0.166374	0.782388	0.130623	-0.106122	-0.073735	1.000000	-0.140636	0.292454	0.083376	0.043961	-0.254428
gdp_growth	-0.192566	-0.112071	-0.145301	-0.013462	0.003324	0.075236	0.020956	-0.048455	-0.002531	-0.015196	-0.035434	0.079293	0.031523	-0.140636	1.000000	-0.128093	0.025588	-0.010859	0.117680
gdp_per_capita	0.135399	0.418459	0.545165	0.017266	0.067525	-0.370648	0.107793	0.231349	0.139674	0.007945	0.667172	-0.184681	0.098209	0.292454	-0.128093	1.000000	0.078850	0.382121	-0.013789
Land Area(Km2)	-0.001629	0.054964	0.074630	0.070355	0.141528	-0.045395	0.627682	0.473280	0.704884	0.044866	0.114661	0.071686	0.653732	0.083376	0.025588	0.078850	1.000000	-0.021608	-0.032837
Latitude	0.003507	0.420548	0.432700	-0.041691	0.021782	-0.330140	0.110994	0.177961	0.112833	-0.029134	0.329893	-0.014179	0.115571	0.043961	-0.010859	0.382121	-0.021608	1.000000	-0.016475
Longitude	0.000944	-0.102857	-0.131847	-0.025587	0.029141	0.018284	0.045413	-0.087111	-0.016380	-0.045336	0.039724	0.135104	0.036286	-0.254428	0.117680	-0.013789	-0.032837	-0.016475	1.000000

Pada dataset ini terdapat 21 variabel yang terdiri dari 1 variabel dependent dan 20 variabel independent. Variabel dependent ini akan dijadikan target dimana variabel ini dipengaruhi oleh variabel lain. Nilai korelasi yang paling tinggi dapat dilihat pada hubungan antara "Value_co2_emissions_kt_by_country dan "Electricity from fossil fuels (TWh) " dengan nilai korelasi sebesar 0.989741. Ini menunjukkan bahwa ada korelasi positif yang kuat antara emisi karbon dioksida dan listrik yang dihasilkan dari bahan bakar fosil.



Data Preparation

Website

www.greatedu.co.id

SIB Cycle 5 |**Data Analyst**

- Mengubah nama fitur

```
#mengganti nama kolom
df = df.rename(columns = {"Density\\n(P/Km2)":"Density (P/km2)",
                  "Value_co2_emissions_kt_by_country":"CO2 emissions value by country (kT)",
                  "Access to clean fuels for cooking":"Access to clean fuels for cooking (% of population)",
                  "gdp_growth":"GDP growth",
                  "gdp_per_capita":"GDP per capita",
                  "Renewable-electricity-generating-capacity-per-capita":"Renewable electricity Generating Capacity per capita"
                  })
```

```
df
```

Entity	Year	Access to electricity (% of population)	Access to clean fuels for cooking (% of population)	Renewable electricity Generating Capacity per capita	Financial flows to developing countries (US $)	Renewable energy share in the total final energy consumption (%)	Electricity from fossil fuels (TWh)	Electricity from nuclear (TWh)	Electricity from renewables (TWh)	...	Primary energy consumption per capita (kWh/person)	Energy intensity level of primary energy (MJ/$2017 PPP GDP)	CO2 emissions value by country (kT)

Mengubah nama fitur atau kolom agar lebih informatif, seragam dan mengatasi masalah seperti karakter khusus yang sulit diakses dalam pengolahan data.

- Mengubah type data

```
#mengkonversi ke tipe data numerik.
df['Density (P/km2)'] = df['Density (P/km2)'].str.replace(',', '')
```

```
#mengubah type data ojc ke tipe data numeric
df['Density (P/km2)'] = pd.to_numeric(df['Density (P/km2)'])
```

Mengubah data type dari object ke float. Pemrosesan ini membantu dalam mempersiapkan data untuk penggunaan lebih lanjut dalam analisis dan mempersiapkan data menjadi format yang lebih sesuai.

- Menghapus Kolom

```
#menghapus fitur yang tidak perlu
df = df.drop(['Financial flows to developing countries (US $)', 'Renewables (% equivalent primary energy)'], axis=1)
```

Menghapus fitur yang memiliki banyak missing value hal ini mencegah terjadinya overfitting.

- Data Cleansing

  Missing value

  Menangani missing value menggunakan nilai median.

```
Entity                                                              0
Year                                                                0
Access to electricity (% of population)                            10
Access to clean fuels for cooking (% of population)               169
Renewable electricity Generating Capacity per capita              931
Financial flows to developing countries (US $)                    2089
Renewable energy share in the total final energy consumption (%)   194
Electricity from fossil fuels (TWh)                                21
Electricity from nuclear (TWh)                                    126
Electricity from renewables (TWh)                                  21
Low-carbon electricity (% electricity)                             42
Primary energy consumption per capita (kWh/person)                  0
Energy intensity level of primary energy (MJ/$2017 PPP GDP)       207
CO2 emissions value by country (kT)                               428
Renewables (% equivalent primary energy)                         2137
GDP growth                                                        317
GDP per capita                                                    282
Density (P/km2)                                                     1
Land Area(Km2)                                                      1
Latitude                                                            1
Longitude                                                           1
dtype: int64
```

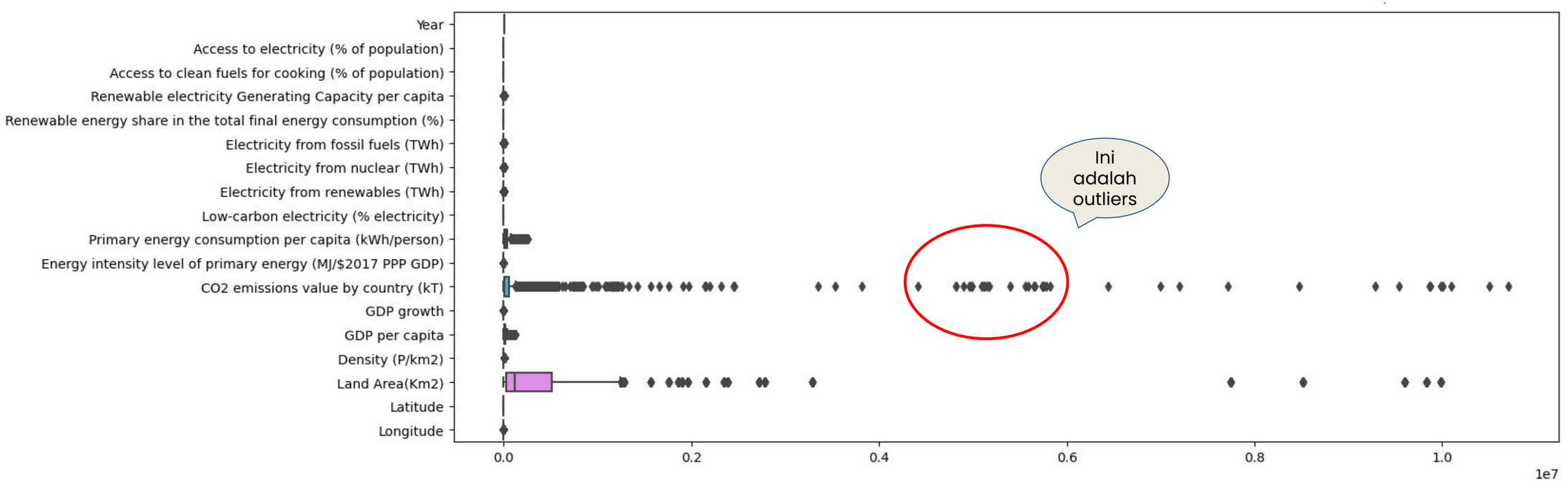
Missing values yang terdeteksi

```
Entity                                                              0
Year                                                                0
Access to electricity (% of population)                             0
Access to clean fuels for cooking (% of population)                 0
Renewable electricity Generating Capacity per capita                0
Renewable energy share in the total final energy consumption (%)    0
Electricity from fossil fuels (TWh)                                 0
Electricity from nuclear (TWh)                                      0
Electricity from renewables (TWh)                                   0
Low-carbon electricity (% electricity)                              0
Primary energy consumption per capita (kWh/person)                  0
Energy intensity level of primary energy (MJ/$2017 PPP GDP)         0
CO2 emissions value by country (kT)                                 0
GDP growth                                                          0
GDP per capita                                                      0
Density (P/km2)                                                     0
Land Area(Km2)                                                      0
Latitude                                                            0
Longitude                                                           0
dtype: int64
```

Jumlah missing values setelah ditangani

Penanganan missing value dapat membantu memastikan bahwa data yang digunakan dalam analisis atau pemodelan konsisten.

Visualisasi Outliers



Outliers

Outliers yang terdeteksi

```
outliers.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2284 entries, 7 to 3648
Data columns (total 19 columns):
 #   Column                                                  Non-Null Count  Dtype
---  ------                                                  --------------  -----
 0   Entity                                                  2284 non-null   object
 1   Year                                                    2284 non-null   int64
 2   Access to electricity (% of population)                 2284 non-null   float64
 3   Access to clean fuels for cooking (% of population)     2284 non-null   float64
 4   Renewable electricity Generating Capacity per capita    2284 non-null   float64
 5   Renewable energy share in the total final energy consumption (%)  2284 non-null   float64
 6   Electricity from fossil fuels (TWh)                     2284 non-null   float64
 7   Electricity from nuclear (TWh)                          2284 non-null   float64
 8   Electricity from renewables (TWh)                       2284 non-null   float64
 9   Low-carbon electricity (% electricity)                  2284 non-null   float64
 10  Primary energy consumption per capita (kWh/person)      2284 non-null   float64
 11  Energy intensity level of primary energy (MJ/$2017 PPP GDP)  2284 non-null   float64
 12  CO2 emissions value by country (kT)                     2284 non-null   float64
 13  GDP growth                                              2284 non-null   float64
 14  GDP per capita                                          2284 non-null   float64
 15  Density (P/km2)                                         2284 non-null   float64
 16  Land Area(Km2)                                          2284 non-null   float64
 17  Latitude                                                2284 non-null   float64
 18  Longitude                                               2284 non-null   float64
dtypes: float64(17), int64(1), object(1)
memory usage: 356.9+ KB
```

Jumlah outliers setelah dibersihkan

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1365 entries, 0 to 3646
Data columns (total 19 columns):
 #   Column                                                  Non-Null Count  Dtype
---  ------                                                  --------------  -----
 0   Entity                                                  1365 non-null   object
 1   Year                                                    1365 non-null   int64
 2   Access to electricity (% of population)                 1365 non-null   float64
 3   Access to clean fuels for cooking (% of population)     1365 non-null   float64
 4   Renewable electricity Generating Capacity per capita    1365 non-null   float64
 5   Renewable energy share in the total final energy consumption (%)  1365 non-null   float64
 6   Electricity from fossil fuels (TWh)                     1365 non-null   float64
 7   Electricity from nuclear (TWh)                          1365 non-null   float64
 8   Electricity from renewables (TWh)                       1365 non-null   float64
 9   Low-carbon electricity (% electricity)                  1365 non-null   float64
 10  Primary energy consumption per capita (kWh/person)      1365 non-null   float64
 11  Energy intensity level of primary energy (MJ/$2017 PPP GDP)  1365 non-null   float64
 12  CO2 emissions value by country (kT)                     1365 non-null   float64
 13  GDP growth                                              1365 non-null   float64
 14  GDP per capita                                          1365 non-null   float64
 15  Density (P/km2)                                         1365 non-null   float64
 16  Land Area(Km2)                                          1365 non-null   float64
 17  Latitude                                                1365 non-null   float64
 18  Longitude                                               1365 non-null   float64
dtypes: float64(17), int64(1), object(1)
memory usage: 213.3+ KB
```

Pembersihan outlier membantu menghasilkan estimasi yang lebih akurat tentang pusat dan sebaran data.

Melakukan Pemilihan fitur

Melakukan pemilihan fitur yang akan digunakan pada pemodelan

```
[ ]  df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1365 entries, 0 to 3646
Data columns (total 16 columns):
 #   Column                                                  Non-Null Count  Dtype
---  ------                                                  --------------  -----
 0   Year                                                    1365 non-null   int64
 1   Access to electricity (% of population)                 1365 non-null   float64
 2   Access to clean fuels for cooking (% of population)     1365 non-null   float64
 3   Renewable electricity Generating Capacity per capita    1365 non-null   float64
 4   Renewable energy share in the total final energy consumption (%)  1365 non-null   float64
 5   Electricity from fossil fuels (TWh)                     1365 non-null   float64
 6   Electricity from renewables (TWh)                       1365 non-null   float64
 7   Low-carbon electricity (% electricity)                  1365 non-null   float64
 8   Primary energy consumption per capita (kWh/person)      1365 non-null   float64
 9   Energy intensity level of primary energy (MJ/$2017 PPP GDP)  1365 non-null   float64
 10  CO2 emissions value by country (kT)                     1365 non-null   float64
 11  GDP growth                                              1365 non-null   float64
 12  GDP per capita                                          1365 non-null   float64
 13  Density (P/km2)                                         1365 non-null   float64
 14  Land Area(Km2)                                          1365 non-null   float64
 15  Longitude                                               1365 non-null   float64
dtypes: float64(15), int64(1)
memory usage: 181.3 KB
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1365 entries, 0 to 3646
Data columns (total 19 columns):
 #   Column                                                  Non-Null Count  Dtype
---  ------                                                  --------------  -----
 0   Entity                                                  1365 non-null   object
 1   Year                                                    1365 non-null   int64
 2   Access to electricity (% of population)                 1365 non-null   float64
 3   Access to clean fuels for cooking (% of population)     1365 non-null   float64
 4   Renewable electricity Generating Capacity per capita    1365 non-null   float64
 5   Renewable energy share in the total final energy consumption (%)  1365 non-null   float64
 6   Electricity from fossil fuels (TWh)                     1365 non-null   float64
 7   Electricity from nuclear (TWh)                          1365 non-null   float64
 8   Electricity from renewables (TWh)                       1365 non-null   float64
 9   Low-carbon electricity (% electricity)                  1365 non-null   float64
 10  Primary energy consumption per capita (kWh/person)      1365 non-null   float64
 11  Energy intensity level of primary energy (MJ/$2017 PPP GDP)  1365 non-null   float64
 12  CO2 emissions value by country (kT)                     1365 non-null   float64
 13  GDP growth                                              1365 non-null   float64
 14  GDP per capita                                          1365 non-null   float64
 15  Density (P/km2)                                         1365 non-null   float64
 16  Land Area(Km2)                                          1365 non-null   float64
 17  Latitude                                                1365 non-null   float64
 18  Longitude                                               1365 non-null   float64
dtypes: float64(17), int64(1), object(1)
memory usage: 213.3+ KB
```

Melakukan Pemilihan fitur

Melakukan pemilihan fitur yang paling relevan dengan target.

```
[ ]  df_cleaned.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1365 entries, 0 to 3646
Data columns (total 8 columns):
 #   Column                                                  Non-Null Count  Dtype
---  ------                                                  --------------  -----
 0   Access to electricity (% of population)                1365 non-null   float64
 1   Access to clean fuels for cooking (% of population)    1365 non-null   float64
 2   Renewable energy share in the total final energy consumption (%)  1365 non-null   float64
 3   Electricity from fossil fuels (TWh)                    1365 non-null   float64
 4   Electricity from renewables (TWh)                      1365 non-null   float64
 5   Primary energy consumption per capita (kWh/person)     1365 non-null   float64
 6   CO2 emissions value by country (kT)                    1365 non-null   float64
 7   GDP per capita                                         1365 non-null   float64
dtypes: float64(8)
memory usage: 96.0 KB
```

```
[ ]  df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1365 entries, 0 to 3646
Data columns (total 16 columns):
 #   Column                                                  Non-Null Count  Dtype
---  ------                                                  --------------  -----
 0   Year                                                    1365 non-null   int64
 1   Access to electricity (% of population)                 1365 non-null   float64
 2   Access to clean fuels for cooking (% of population)     1365 non-null   float64
 3   Renewable electricity Generating Capacity per capita    1365 non-null   float64
 4   Renewable energy share in the total final energy consumption (%)  1365 non-null   float64
 5   Electricity from fossil fuels (TWh)                     1365 non-null   float64
 6   Electricity from renewables (TWh)                       1365 non-null   float64
 7   Low-carbon electricity (% electricity)                  1365 non-null   float64
 8   Primary energy consumption per capita (kWh/person)      1365 non-null   float64
 9   Energy intensity level of primary energy (MJ/$2017 PPP GDP)  1365 non-null   float64
 10  CO2 emissions value by country (kT)                     1365 non-null   float64
 11  GDP growth                                              1365 non-null   float64
 12  GDP per capita                                          1365 non-null   float64
 13  Density (P/km2)                                         1365 non-null   float64
 14  Land Area(Km2)                                          1365 non-null   float64
 15  Longitude                                               1365 non-null   float64
dtypes: float64(15), int64(1)
memory usage: 181.3 KB
```

Memisahkan Target dari DataFrame

```
X
```

	Access to electricity (% of population)	Access to clean fuels for cooking (% of population)	Renewable energy share in the total final energy consumption (%)	Electricity from renewables (TWh)	Primary energy consumption per capita (kWh/person)	CO2 emissions value by country (kT)	GDP per capita
0	-2.198293	-1.148345	0.207224	-0.534350	-0.807848	-0.641664	-0.698087
1	-2.119203	-1.122876	0.227671	-0.483990	-0.812621	-0.643014	-0.698087
2	-1.947762	-1.097408	-0.032780	-0.468086	-0.814512	-0.629519	-0.671067
3	-1.776490	-1.064298	-0.071998	-0.449532	-0.813124	-0.620972	-0.669353
4	-1.605310	-1.028642	0.182084	-0.468086	-0.814994	-0.629519	-0.666201
...	...	...	...	...	...	...	...
1360	-0.819241	-1.041376	-0.519826	2.855697	-0.604689	-0.203535	-0.548370
1361	-1.168563	-0.453044	1.020755	0.295271	-0.472439	-0.059591	-0.612646
1362	-1.151917	-0.450497	1.096845	0.244910	-0.494565	-0.050595	-0.611834
1363	-0.830356	-0.547279	1.463219	0.523218	-0.606947	-0.210732	-0.510290
1364	-0.785556	-0.544732	1.388469	0.830681	-0.579778	-0.118968	-0.507327

1365 rows × 7 columns

```
y
```

```
0       0.16
1       0.09
2       0.13
3       0.31
4       0.33
        ...
3627    2.00
3628    3.30
3629    4.33
3645    3.05
3646    3.73
Name: Electricity from fossil fuels (TWh), Length: 1365, dtype: float64
```

```
# Memisahkan kolom target
X = df_cleaned.drop('Electricity from fossil fuels (TWh)', axis=1)
y = df_cleaned['Electricity from fossil fuels (TWh)'] # target
```

```
data_scaler= StandardScaler()
X_scaled= data_scaler.fit_transform(X)
X= pd.DataFrame(X_scaled,columns=X.columns)
```

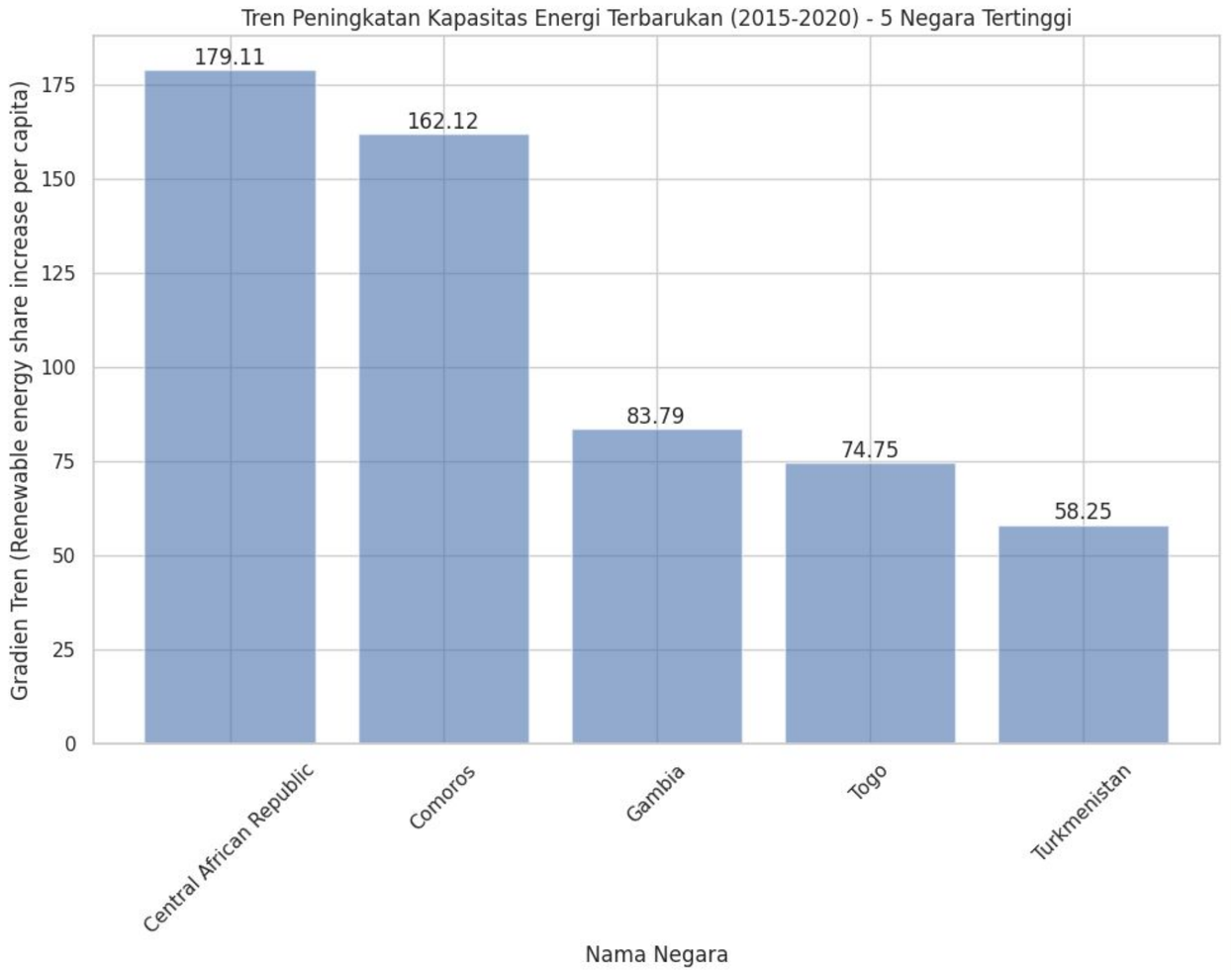
Melakukan pemodelan dengan mendrop kolom target dan melakukan standardscaler.

EDA (Exploratory Data Analysis)

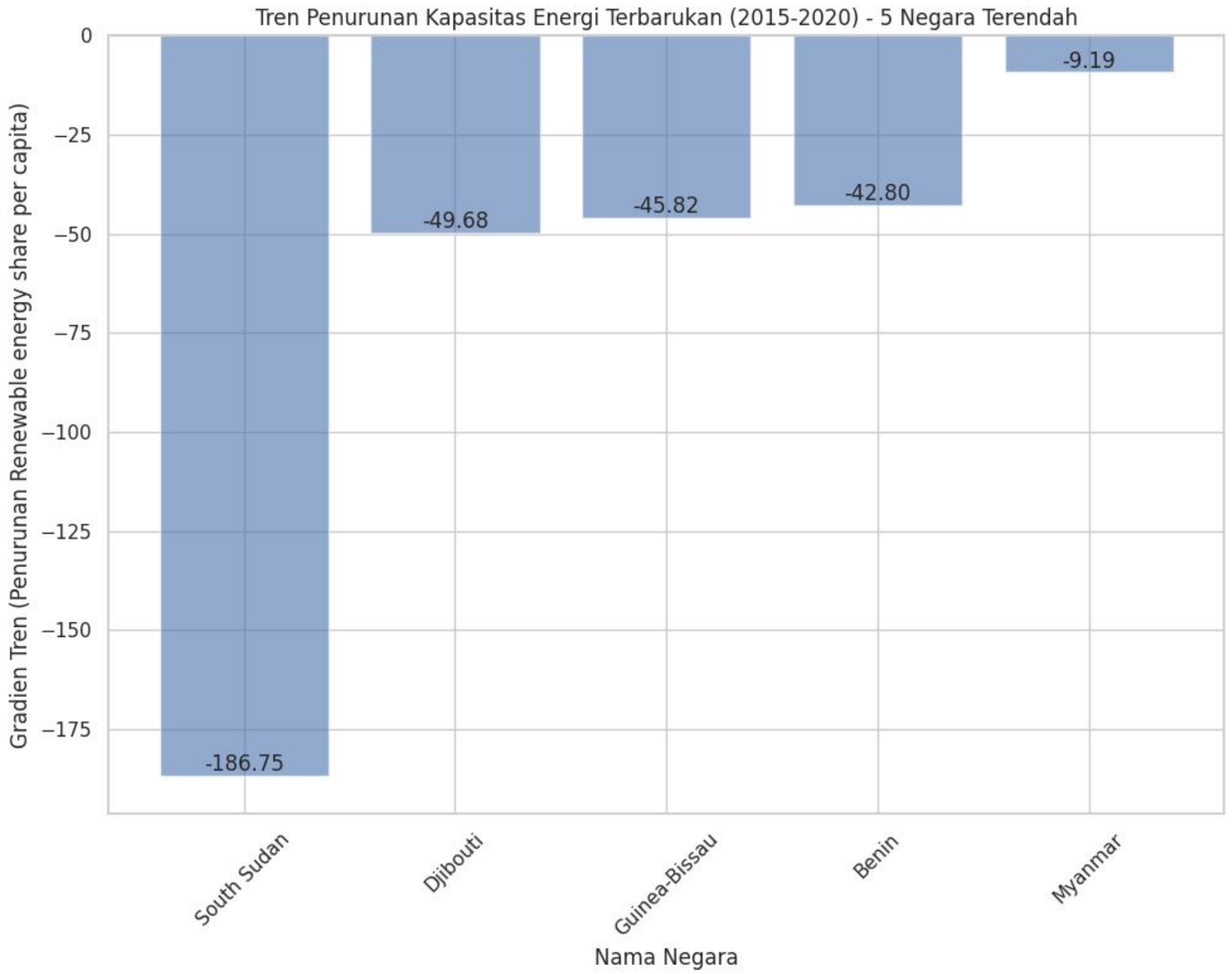
Swipe

Pertumbuhan Kapasitas Energi Terbarukan pada Lima Negara Tertinggi (2015-2020)

Berdasarkan Histogram di samping negara yang menunjukkan peningkatan energi terbarukan adalah Central African Republic dengan peningkatan selama 5 tahun terakhir sebesar 179,11.



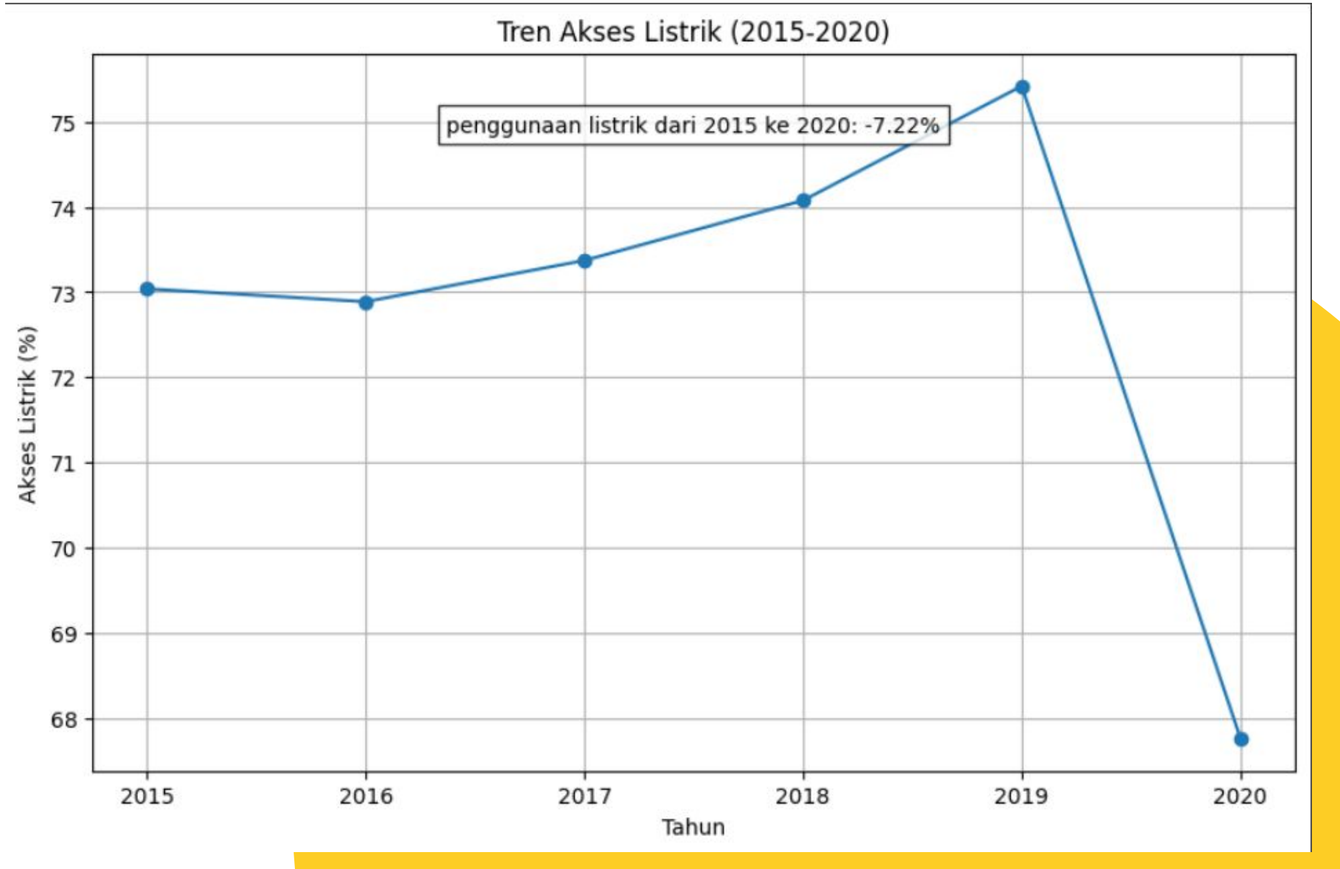
Penurunan Kapasitas Energi Terbarukan (2015-2020) Pada Lima Negara dengan Tren Terendah



Berdasarkan Histogram di samping negara yang menunjukkan penurunan energi terbarukan adalah South Sudan dengan penurunan selama 5 tahun terakhir sebesar 186,75.

Tren Pengaksesan Listrik Selama 5 tahun terakhir

Grafik disamping menunjukkan pengaksesan listrik cenderung meningkat. Namun pada tahun 2020 terjadi penurunan sebanyak 7,22%.

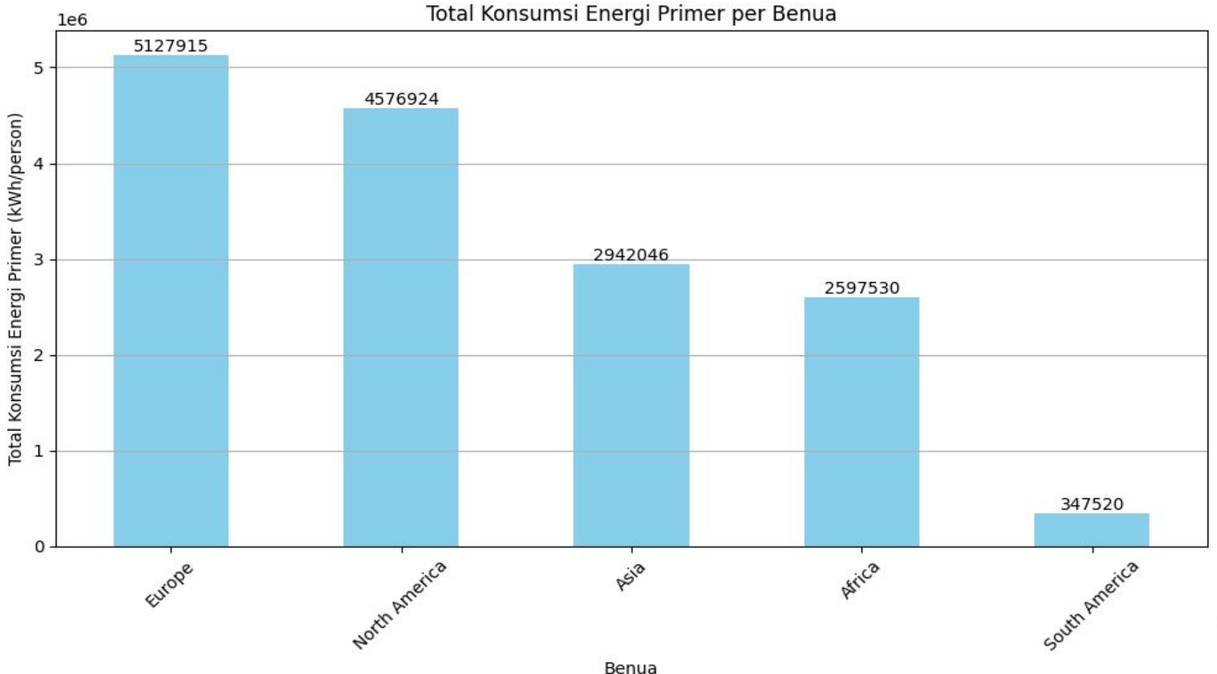


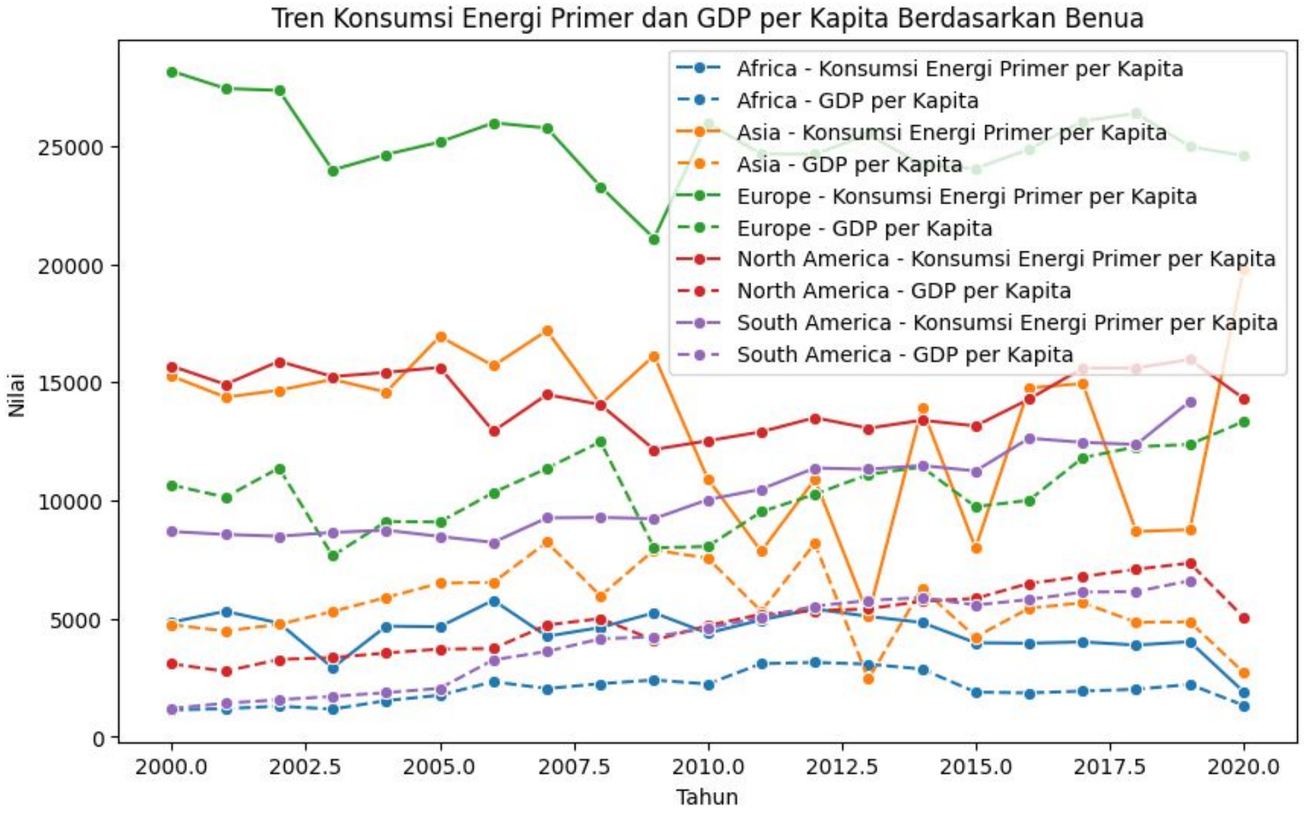
Tren Penggunaan bahan bakar ramah lingkungan

Penggunaan bahan bakar yang ramah lingkungan cenderung tidak stabil dari tahun 2000 sampai 2020. Untuk penggunaan bahan bakar ramah lingkungan yang paling tinggi terjadi pada tahun 2002 dan 2012. Selain itu, terjadi penurunan yang cukup ekstrim pada tahun 2009..

Tren Konsumsi Energi Primer dan GDP per Kapita Berdasarkan Benua dari Tahun ke Tahun

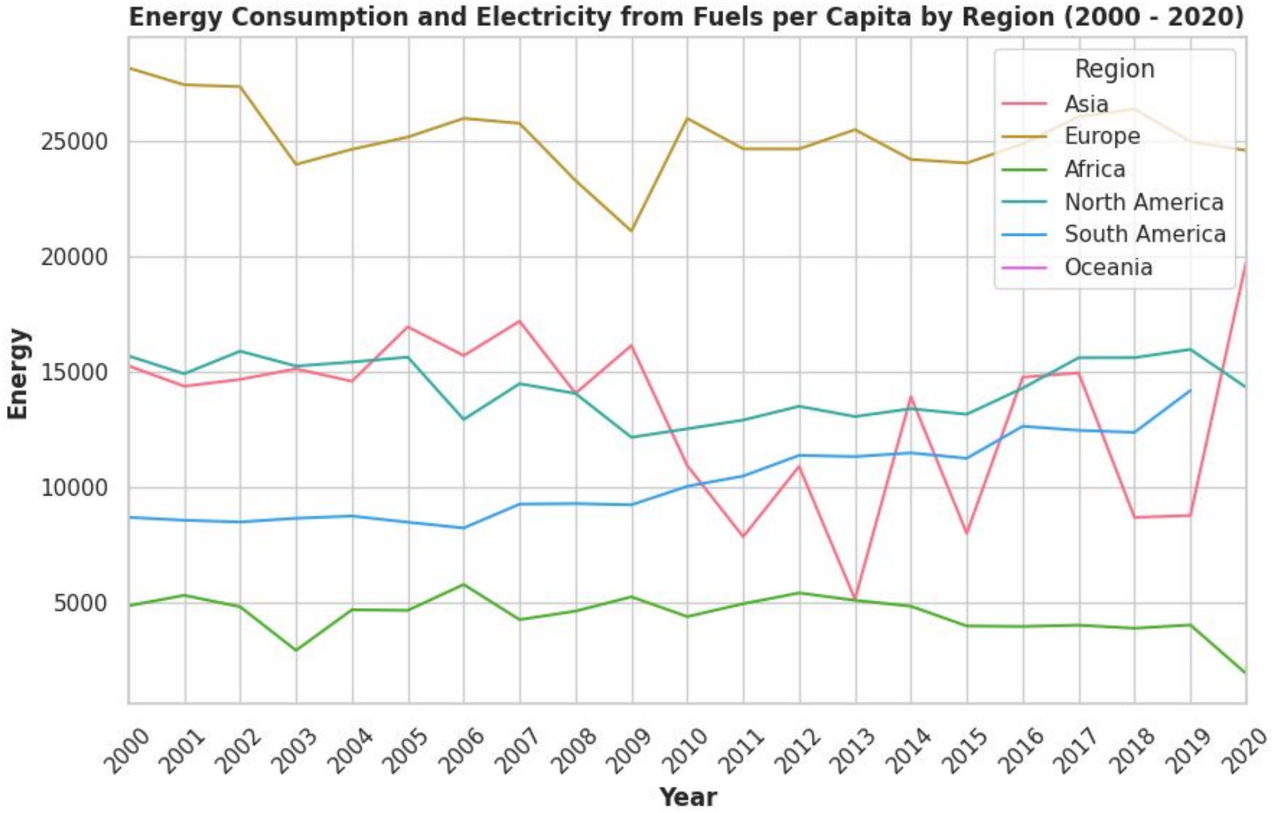
Visualisasi disamping menggambarkan tren yang membandingkan konsumsi energi primer per kapita dan GDP (Gross Domestic Product) per kapita dari berbagai benua. Berdasarkan grafik disamping benua yang paling banyak menggunakan primer yaitu benua Eropa dengan GDP benua tersebut yang paling besar dibandingkan benua lainnya.





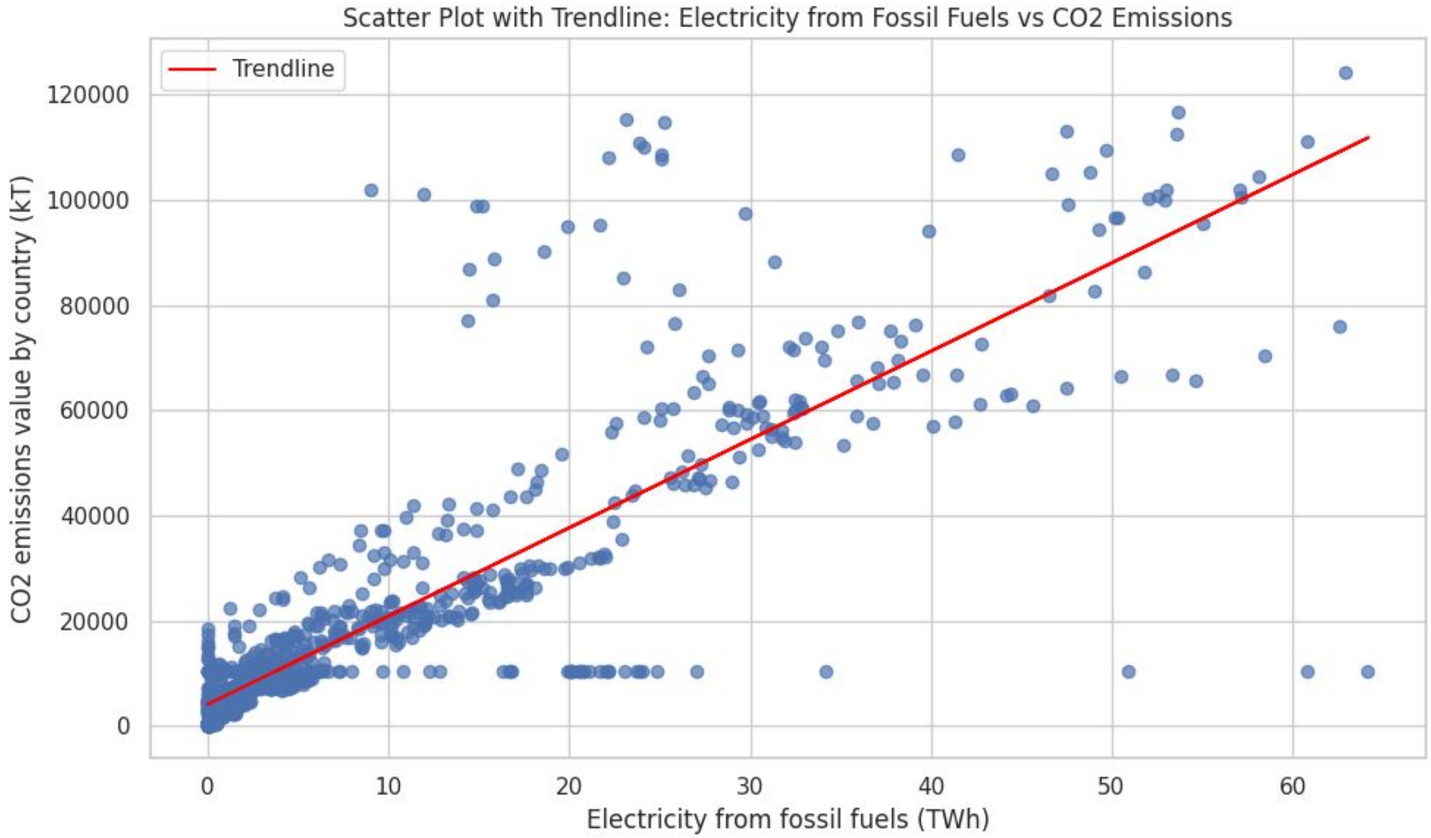
Perubahan Konsumsi Energi dan Listrik dari Bahan Bakar Fosil per Kapita Menurut Wilayah (2000-2020)

Pada grafik disamping setiap tahun dari masing masing benua mengalami perubahan konsumsi energi dan listrik dari bahan bakar fosil. Pada satu tahun terakhir setiap benua mengalami penurunan. Namun, untuk benua Asia dan South America mengalami kenaikan dengan Asia yang menduduki kenaikan tertinggi.



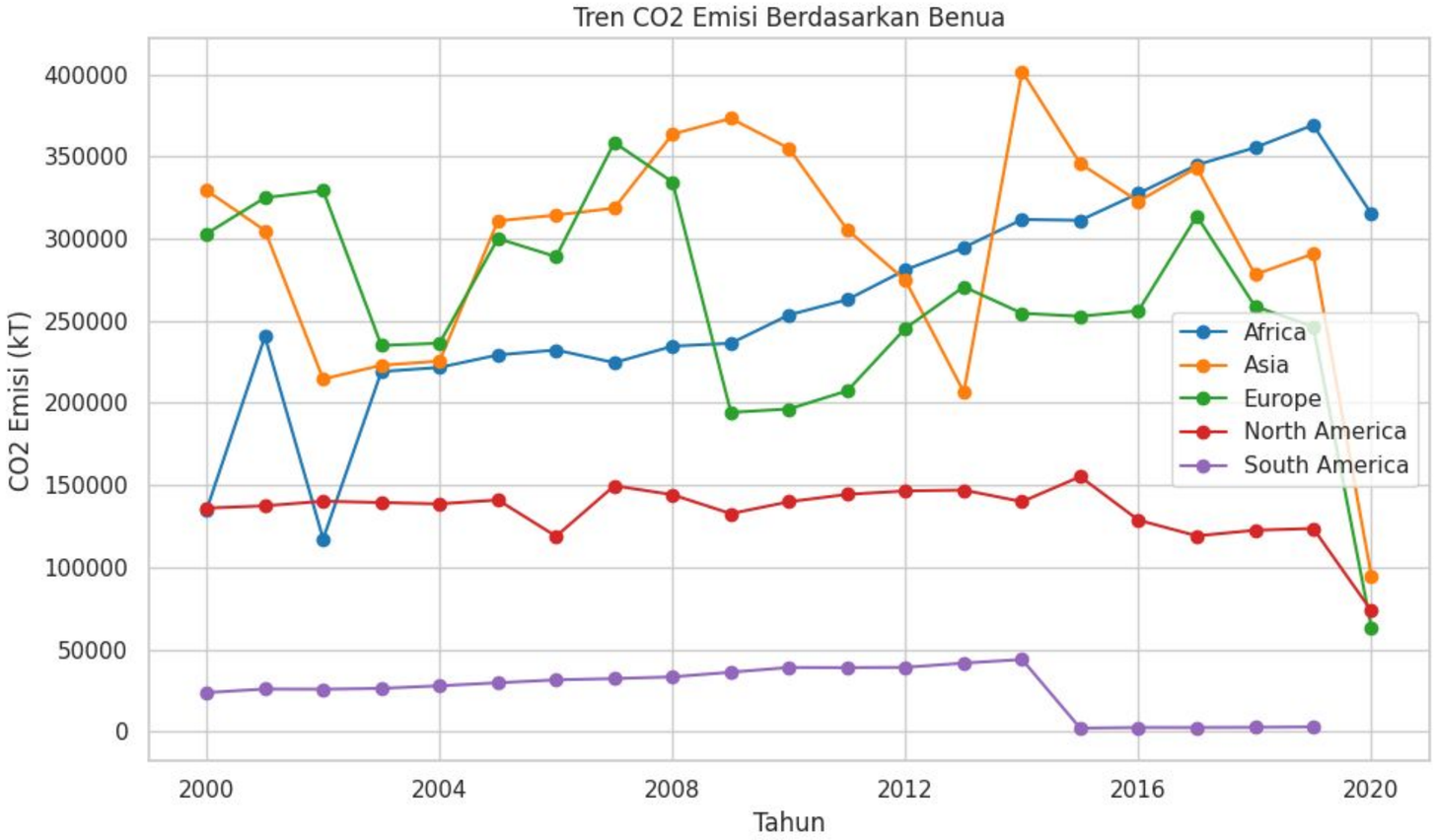
Hubungan Antara Konsumsi Listrik dari Bahan Bakar Fosil dan Emisi CO2

Berdasarkan scatter plot dan juga garis trendline yang naik menunjukkan hubungan positif di antara kedua variabel, yang berarti semakin tinggi konsumsi listrik dari bahan bakar fosil, semakin tinggi juga emisi CO2.



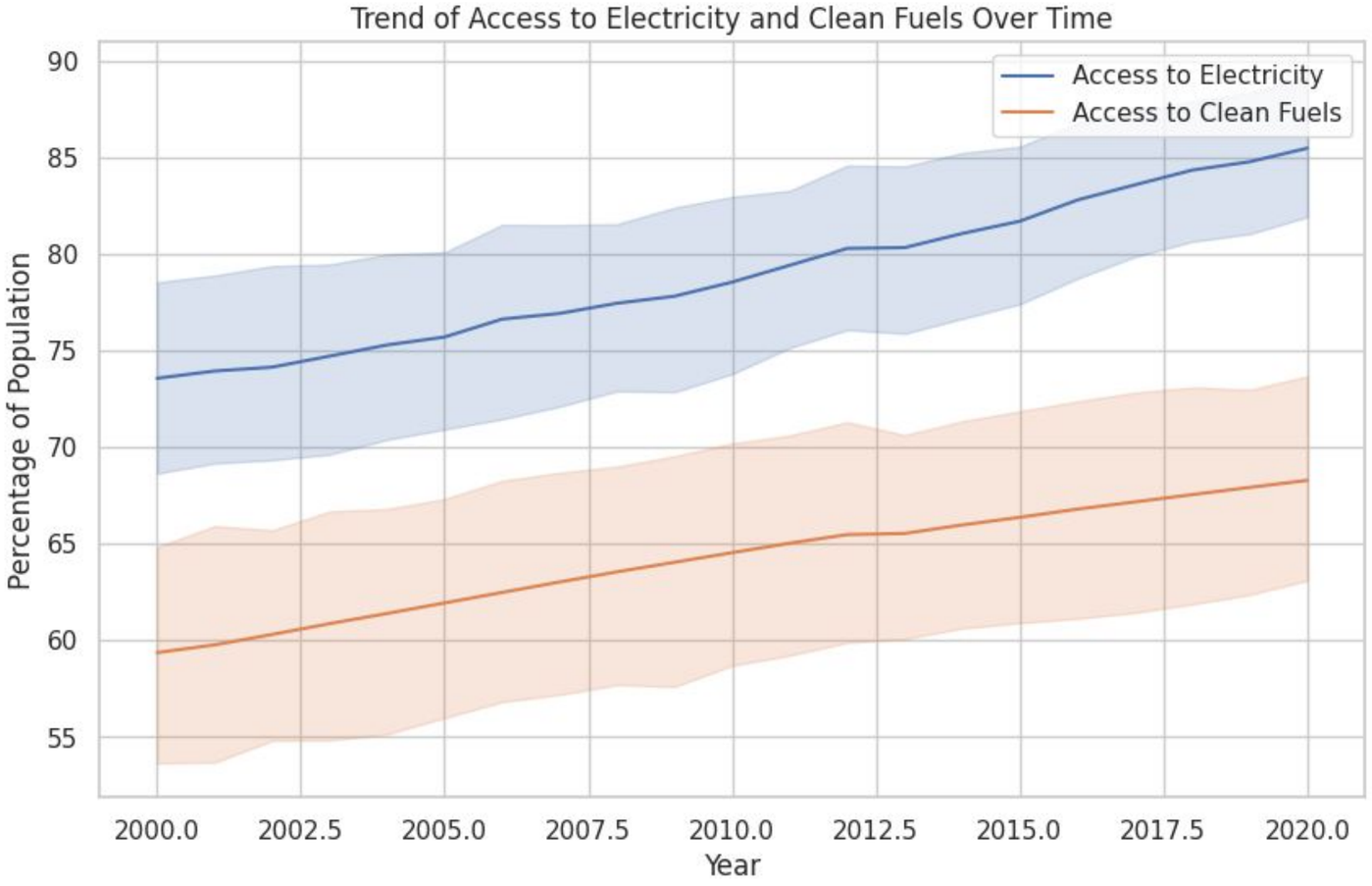
Tren Emisi CO2 Berdasarkan Benua dari Tahun ke Tahun

Grafik disamping menunjukkan perubahan emisi CO2 dari setiap benua. Kenaikan paling tinggi terjadi pada benua Asia yang mencapai 400.000 kT yang terjadi pada rentang tahun 2012 sampai 2016. Pada tahun 2020 setiap benua mengalami penurun emisi CO2

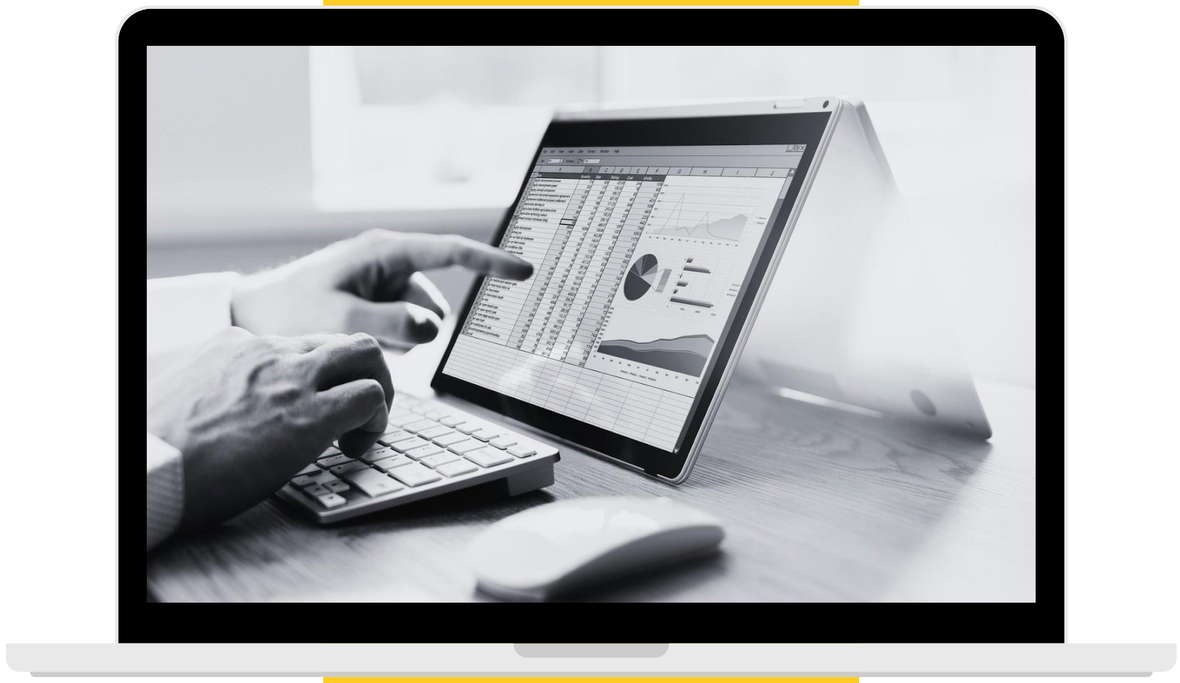


Perbandingan Akses Listrik dan Penduduk yang bergantung Pada Bahan Bakar Ramah Lingkungan

Pada tahun 2000, sekitar 70% penduduk memiliki akses listrik dan angka ini naik menjadi hampir 90% pada tahun 2020. Di tahun 2020, sekitar 65% penduduk masih bergantung pada bahan bakar ramah lingkungan.



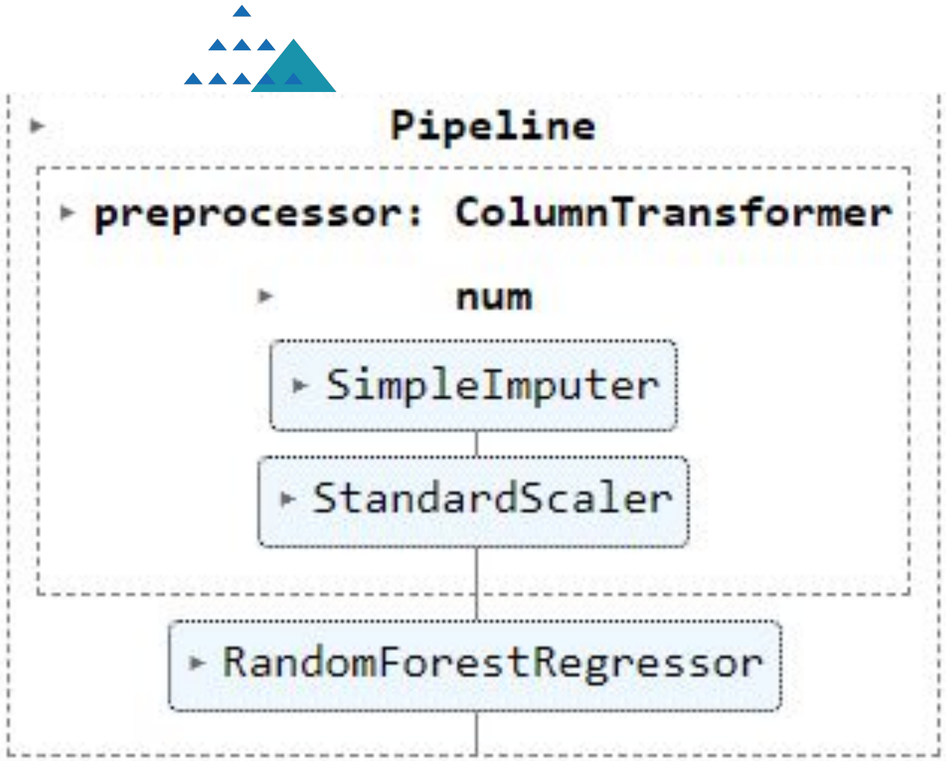
Modelling

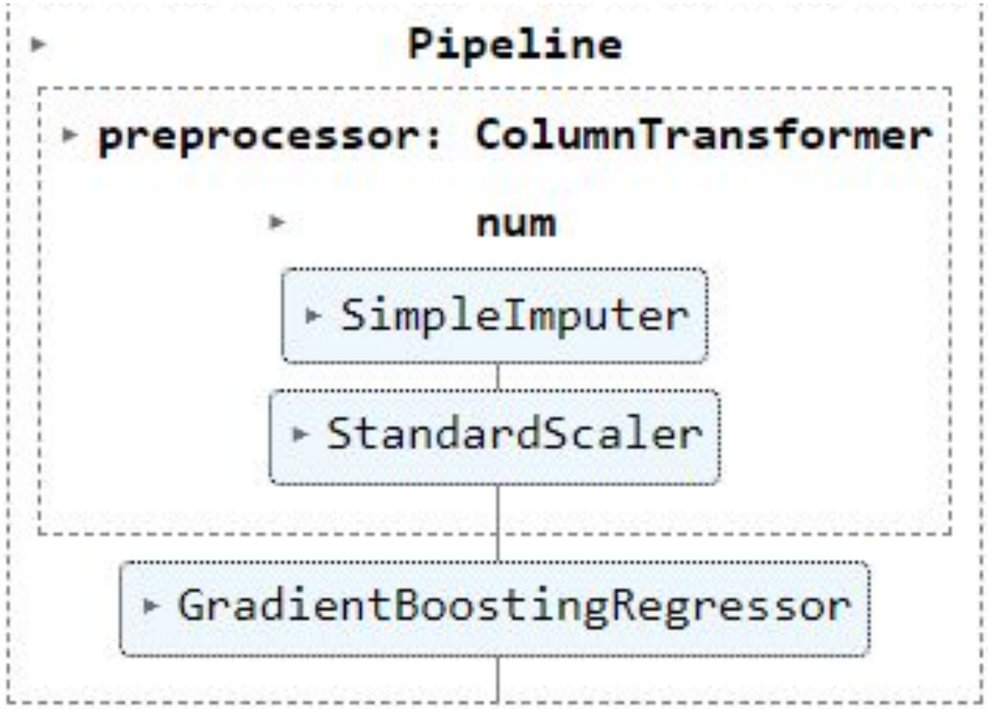


Pemodelan

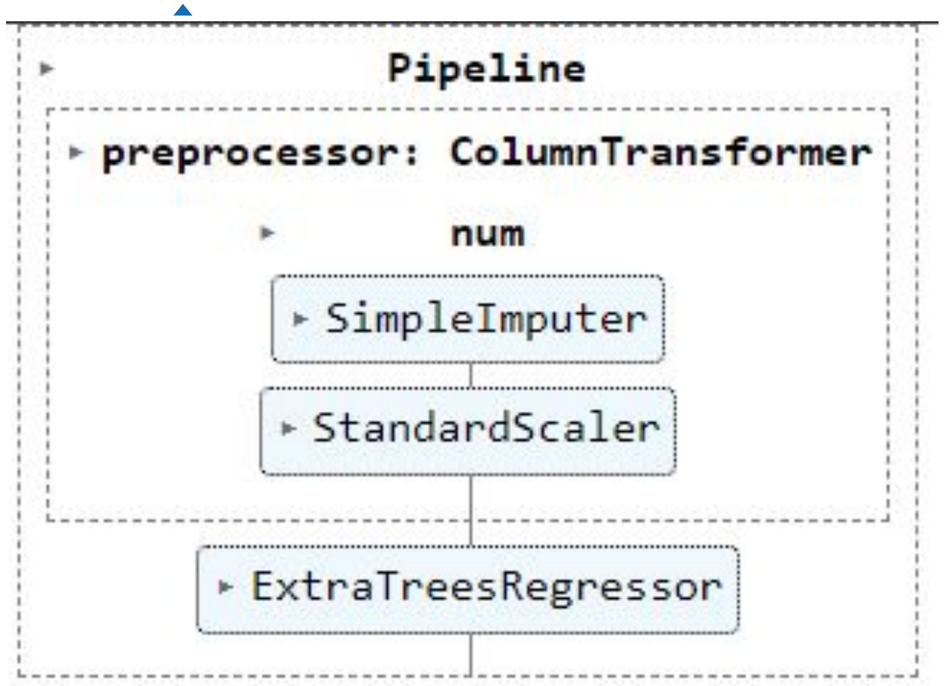
Pemodelan yang digunakan adalah pemodelan supervised learning yaitu regresi dengan kolom targetnya adalah Electricity From Fuels. Pemodelan ini menggunakan Decision Tree Regressor, Random Forest Regressor, Extra Tree Regressor dan Gradient Boosting Regressor.

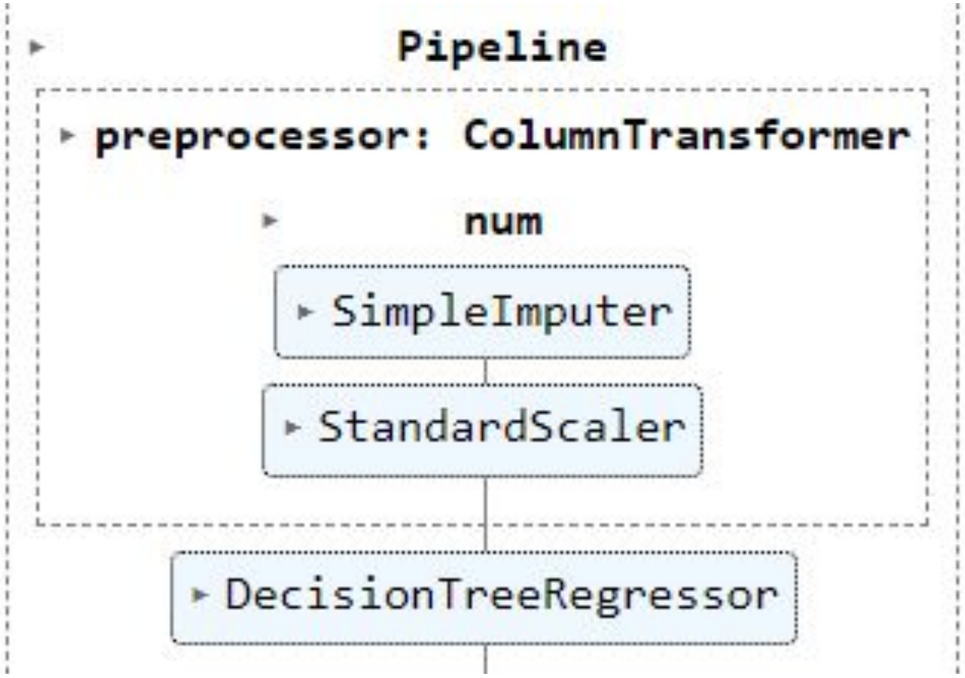
Melakukan Pemodelan





Melakukan Pemodelan





Hasil Prediksi

Nilai predict

Nilai aktual

	RandomForest_Prediction	GradientBoosting_Prediction	ExtraTrees_Prediction	DecisionTree_Prediction	Electricity from fossil fuels (TWh)
0	32.46565	31.387824	33.374100	31.931667	30.84
1	24.94785	24.460415	24.231742	29.756250	24.13
2	0.03925	0.035119	0.040363	0.042727	0.04
3	0.86055	1.040040	0.882242	0.646210	0.92
4	1.83535	1.318827	1.729092	1.532976	2.19
...	...	...	...	...	...
268	0.09405	0.110455	0.093283	1.532976	0.10
269	33.86140	28.194367	29.018454	37.693529	32.38
270	14.41080	13.813846	14.157000	15.172857	14.13
271	2.97325	2.307932	3.013196	1.532976	2.58
272	0.30645	0.196384	0.311167	0.157829	0.32

273 rows × 5 columns

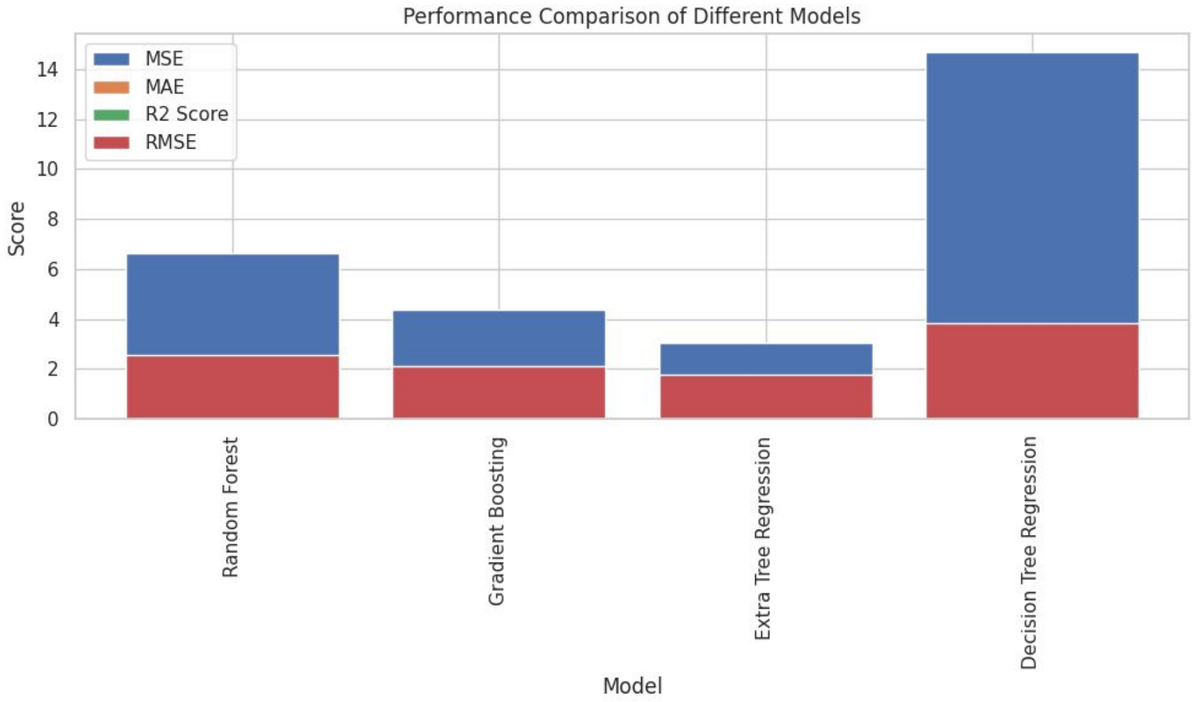
Evaluasi

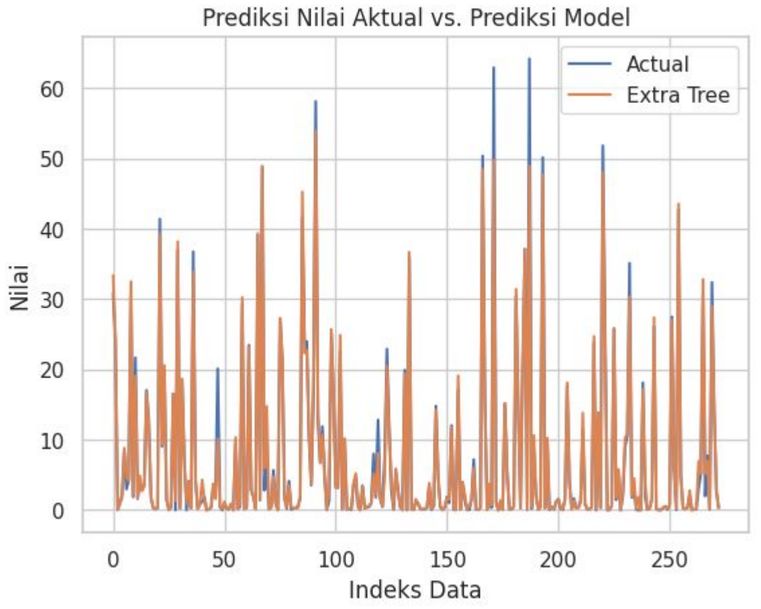
Swipe

Matriks Evaluasi

	Model	MSE	MAE	R2 Score	RMSE
0	Random Forest	6.623396	0.810628	0.957529	2.573596
1	Gradient Boosting	4.386702	0.875327	0.971871	2.094446
2	Extra Tree Regression	3.064039	0.686837	0.980352	1.750440
3	Decision Tree Regression	14.672345	1.441283	0.905916	3.830450

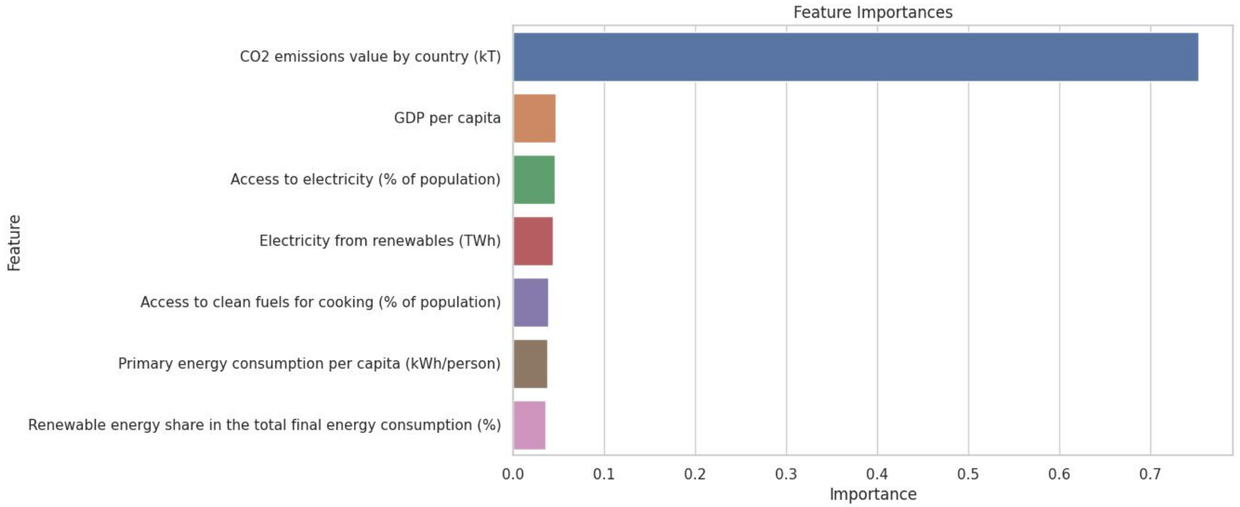
Berdasarkan Hasil matriks evaluasi model Extra tree Regressor, pemodelan yang paling baik dengan skor RMSE sebesar 1,750440.



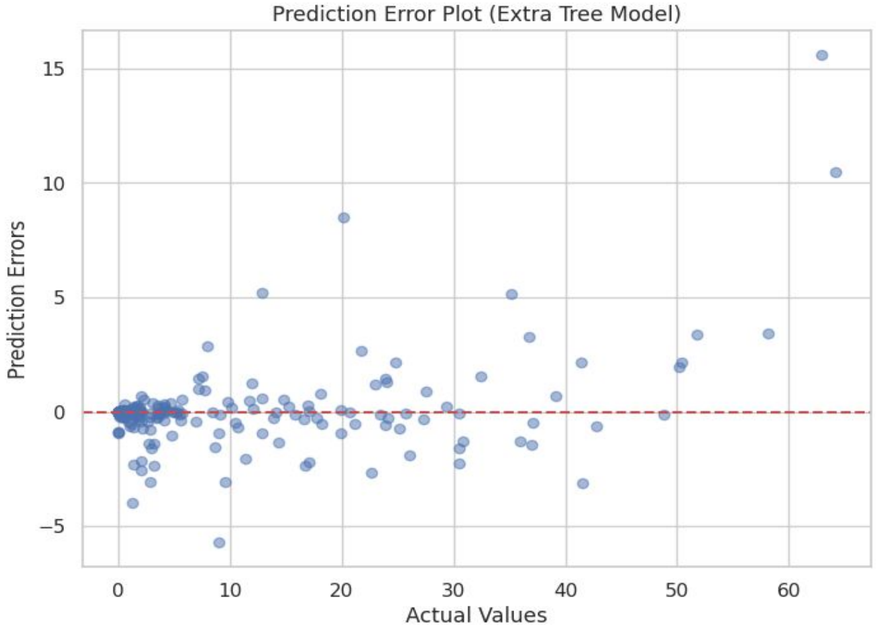


www.greatedu.co.id

Feature Importance

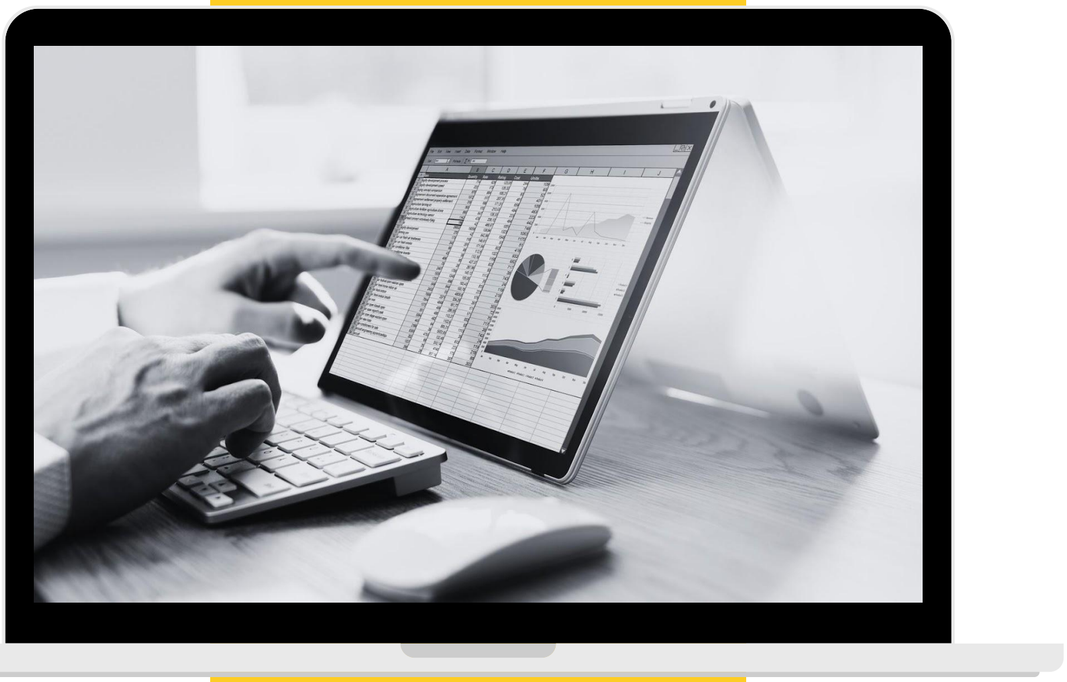


	Model	MSE	MAE	R2 Score	RMSE
0	Random Forest	6.623396	0.810628	0.957529	2.573596
1	Gradient Boosting	4.386702	0.875327	0.971871	2.094446
2	Extra Tree Regression	3.064039	0.686837	0.980352	1.750440
3	Decision Tree Regression	14.672345	1.441283	0.905916	3.830450



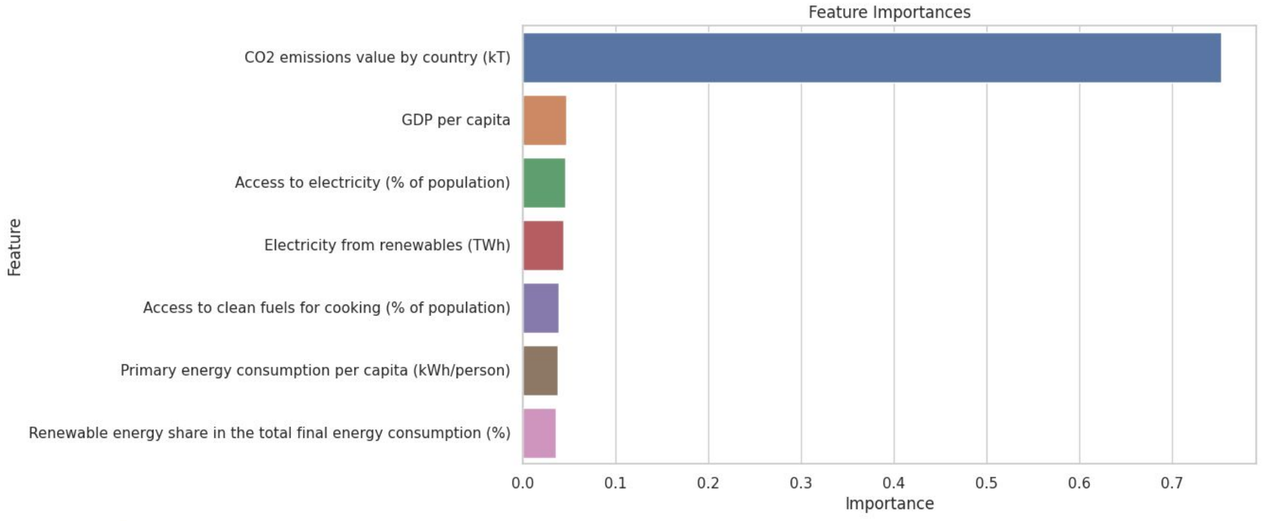
Kesimpulan

Berdasarkan pemodelan yang sudah dilakukan pemodelan yang terbaik adalah Extra Tree Regression dengan skor RMSE sebesar 1,750440 dimana skor ini yang terbaik karena mendekati 0.



Deployment

Feature Importance



Deployment

Hasil proyek ini adalah laporan nilai produksi listrik dari bahan bakar fosil (batu bara, minyak, gas) dalam satuan terawatt-hour (TWh), dengan tujuan untuk memprediksi produksi listrik dari bahan bakar fosil. Berdasarkan analisis regresi, terbukti bahwa emisi CO2 menjadi faktor utama yang mempengaruhi produksi listrik tersebut, dengan tingkat signifikansi sebesar 0.753765. Mengurangi emisi CO2 sangat penting sejalan dengan Sustainable Development Goals (SDGs), terutama SDG 7, 13, dan 12. Untuk mencapai target-target SDGs tersebut, maka harus dilakukan pengembangan sumber energi berkelanjutan dan peningkatan efisiensi energi agar dapat mengurangi ketergantungan pada bahan bakar fosil.

Terima kasih