

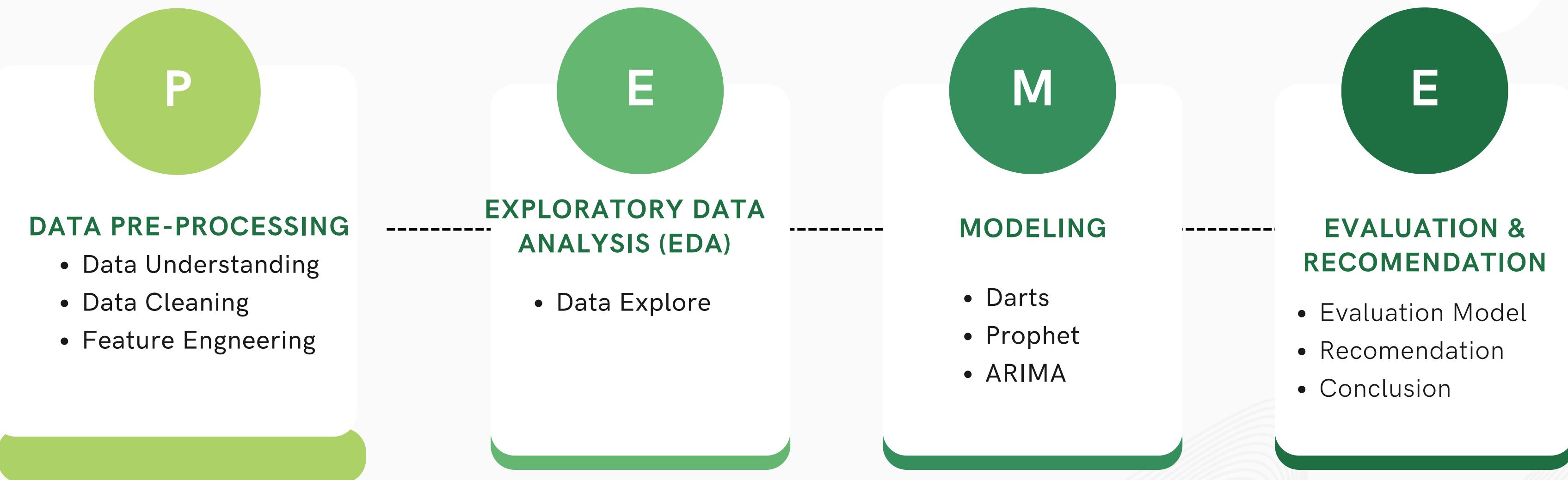
| Final Project

# Predicting National Energy Consumption: A Comprehensive Multi-Forecasting Approach

**Muhammad Arkhan Doohan**

Dibimbing (DS25)

# Outline Presentation



## | Data Pre-Processing

The sustainable product market  
is growing rapidly

---



# | Data Understanding

## "Pahami maksud dari data"

Bagian ini saya mencoba memahami apa saja maksud dari setiap kolom atas dataset yang saya gunakan. Gambaran besar dari dataset ini yaitu penggunaan berbagai energi dari setiap negara yang terjadi dalam kurun waktu 20 tahun (2000-2020).

Penjelasan Umum

Key Features:

- Entity: Nama negara atau wilayah tempat data dilaporkan.
- Year: Tahun dimana data dilaporkan, berkisar antara tahun 2000 hingga 2020.
- Access to electricity (% of population): Persentase penduduk yang memiliki akses terhadap listrik.
- Access to clean fuels for cooking (% of population): Persentase penduduk yang sangat bergantung pada bahan bakar ramah lingkungan.
- Renewable-electricity-generating-capacity-per-capita: Kapasitas energi terbarukan terpasang per orang
- Financial flows to developing countries (US\$): Aid and assistance Bantuan dan bantuan dari negara maju untuk proyek energi bersih.
- Electricity from fossil fuels (TWh): Listrik dihasilkan dari bahan bakar fosil (batubara, minyak, gas) dalam terawatt-jam.
- Electricity from nuclear (TWh): Listrik yang dihasilkan dari tenaga nuklir dalam terawatt-jam.
- Electricity from renewables (TWh): Listrik dihasilkan dari sumber terbarukan (hidro, surya, angin, dll.) dalam terawatt-jam.
- Low-carbon electricity (% electricity): Persentase listrik dari sumber rendah karbon (nuklir dan energi terbarukan).
- Primary energy consumption per capita (kWh/person): Konsumsi energi per orang dalam kilowatt-jam.
- Energy intensity level of primary energy (MJ/\$2011 PPP GDP): Penggunaan energi per unit PDB pada paritas daya beli.
- Value\_co2\_emissions (metric tons per capita): Emisi karbon dioksida per orang dalam metrik ton
- Renewables (% equivalent primary energy): Energi primer setara yang berasal dari sumber terbarukan.
- GDP growth (annual %): Tingkat pertumbuhan PDB tahunan berdasarkan mata uang lokal yang konstan.
- GDP per capita: Produk domestik bruto per orang.

Key Question : Memprediksi penggunaan emisi yang dilakukan oleh setiap negara

Tujuan : Project ini diharapkan mampu mempersiapkan strategi bagi setiap negara untuk lebih mencari alternatif atas kondisi kebutuhan energi tersebut.

Data Pre-Processing

EDA

Modeling

Recomendation

# Load Dataset

## “Menggunakan Gdrive”

Melakukan load data bisa dari berbagai sumber, namun metode yang saya gunakan yaitu melalui google drive (drive.mount) dengan memasukan url directory tempat data tersebut tersimpan.

Setelah data terbaca, maka akan terlihat gambaran dari data yang sudah kita buat, untuk lebih jelasnya saya menggunakan prompt info, shape dan sum.isnull() untuk mengetahui apakah data ini terdapat missing value, terdapat berapa baris dan kolom, serta tipe data apa saja yang ada di dalam data kita.

```
✓ Load Dataset

[ ] import pandas as pd

▶ from google.colab import drive
drive.mount('/content/drive')

⚡ Mounted at /content/drive

[ ] df = pd.read_csv('/content/drive/My Drive/Dibimbing/FinalProject/global-data-on-sustainable-energy.csv')
```

	Entity	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)	Primary energy consumption (kWh/person)	Primary energy intensity (MJ/\$2017 PPP GDP)
0	Afghanistan	2000	1.613591	6.2	9.22	20000.0	44.99	0.16	0.0	0.31	...	302.59482 1.64
1	Afghanistan	2001	4.074574	7.2	8.86	130000.0	45.60	0.09	0.0	0.50	...	236.89185 1.74
2	Afghanistan	2002	9.409158	8.2	8.47	3950000.0	37.83	0.13	0.0	0.56	...	210.86215 1.40
3	Afghanistan	2003	14.738506	9.5	8.09	25970000.0	36.66	0.31	0.0	0.63	...	229.96822 1.40
4	Afghanistan	2004	20.064968	10.9	7.75	NaN	44.24	0.33	0.0	0.56	...	204.23125 1.20
...	...	...	...	...	...	...	...	...	...	...	...	...
3644	Zimbabwe	2016	42.561730	29.8	62.88	30000.0	81.90	3.50	0.0	3.32	...	3227.68020 10.00
3645	Zimbabwe	2017	44.178635	29.8	62.33	5570000.0	82.46	3.05	0.0	4.30	...	3068.01150 9.51
3646	Zimbabwe	2018	45.572647	29.9	82.53	10000.0	80.23	3.73	0.0	5.46	...	3441.98580 9.83

```
✓ Data Pre Prosessing

[ ] df
⚡ Show hidden output
▶ df.info()

◀ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 3649 entries, 0 to 3648
Data columns (total 21 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Entity          3649 non-null   object 
 1   Year            3649 non-null   int64  
 2   Access to electricity (% of population) 3639 non-null   float64
 3   Access to clean fuels for cooking        3480 non-null   float64
 4   Renewable-electricity-generating-capacity-per-capita 2718 non-null   float64
 5   Financial flows to developing countries (US $) 1560 non-null   float64
 6   Renewable energy share in the total final energy consumption (%) 3455 non-null   float64
 7   Electricity from fossil fuels (TWh)    3628 non-null   float64
 8   Electricity from nuclear (TWh)       3523 non-null   float64
 9   Electricity from renewables (TWh)     3628 non-null   float64
 10  Low-carbon electricity (% electricity) 3607 non-null   float64
 11  Primary energy consumption per capita (kWh/person) 3649 non-null   float64
 12  Energy intensity level of primary energy (MJ/$2017 PPP GDP) 3442 non-null   float64
 13  Value_co2_emissions_kt_by_country      3221 non-null   float64
 14  Renewables (% equivalent primary energy) 1512 non-null   float64
 15  gdp_growth                     3332 non-null   float64
 16  gdp_per_capita                  3367 non-null   float64
 17  Density(n/P/km2)                3648 non-null   object  
 18  Land Area(km2)                  3648 non-null   float64
 19  Latitude                       3648 non-null   float64
```

Data Pre-Processing

EDA

Modeling

Recomendation

# Data Cleaning

"Bukan hanya menghapus melainkan mempertimbangkan makna di dalamnya"

Setelah dilakukan pengecekan, data tersebut memiliki banyak sekali missing value. Saya membagi menjadi 3 kategori: hapus kolom(jika missing value sudah terlalu banyak melebihi 50%), hapus baris (jika missing value nya sangat sedikit sekali), input missing value(dengan rata-rata atau median).

## \*Catatan:

Perlu diperhatikan pada saat nanti di dunia profesional seorang DS/DA wajib bertanya terlebih dahulu kepada atasan atau orang yang lebih paham terkait bisnis ketika adanya missing value. Penggunaan di atas hanya metode umum saja, yang biasa dilakukan DS/DA dalam menghadapi missing value.

## \*Input: Missing Value

Sehubung sisa missing value itu memiliki banyak outlier, maka data tersebut saya masukan median.(sebaliknya akan diisi mean jika outliernya tidak ada)

Before

```
[ ] df.isnull().sum() #cek missing value
```

Entity	Year	0
Access to electricity (% of population)		0
Access to clean fuels for cooking		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
Value_co2_emissions_kt_by_country		428
Renewables (% equivalent primary energy)		2137
gdp_growth		317
gdp_per_capita		282
Density\n(P/Km2)		1
Land Area(Km2)		1
Latitude		1
Longitude		1
dtype: int64		

After

```
▶ # Menghapus baris yang memiliki missing value
df.dropna(inplace=True)

# Verifikasi apakah masih ada missing value
print(df.isnull().sum())
```

Entity	Year	0
Access to electricity (% of population)		0
Access to clean fuels for cooking		0
Renewable-electricity-generating-capacity-per-capita		0
Financial flows to developing countries (US \$)		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
Value_co2_emissions_kt_by_country		0
Renewables (% equivalent primary energy)		0
gdp_growth		0
gdp_per_capita		0
Density\n(P/Km2)		0
Land Area(Km2)		0
Latitude		0
Longitude		0
dtype: int64		

Data Pre-Processing

EDA

Modeling

Recomendation

# | Exploratory Data Analysis (EDA)

**Emphasizing affection and appreciation for the natural world.**

---



# Exploratory Data Analysis (EDA)

## "Analisa Insight di dalam data"

Terdapat beberapa poin yang berhasil dikulik setelah melakukan EDA:

1. Jumlah total negara yang ada di dalam data yaitu 175 Negara, dengan terdapat 3 negara yang tidak lengkap selama 20 tahun, yaitu : Montenegro, Serbia dan South Sudan.
2. Negara-negara yang memiliki memiliki emisi tertinggi dari berbagai sektor energi nuklir(amerika), fosil (China), Renewable (China).

Top 10 Negara dengan Emisi Nuklir Tertinggi:	
Entity	
United States	16602.61
France	8914.20
Japan	3386.87
China	2609.19
Germany	2602.75
Canada	1844.24
Ukraine	1783.11
United Kingdom	1493.76
Sweden	1347.55
Spain	1240.44

Name: Electricity from nuclear (TWh), dtype: float64

Top 10 Negara dengan Emisi Fosil Tertinggi:	
Entity	
China	67927.33
United States	57652.00
India	17445.83
Japan	15140.82
Germany	7640.41
Saudi Arabia	5188.39
United Kingdom	5131.52
South Africa	4472.77
Italy	4402.56
Mexico	4303.04

Name: Electricity from fossil fuels (TWh), dtype: float64

Top 10 Negara dengan Emisi Energi Terbarukan Tertinggi:	
Entity	
China	19702.94
United States	10234.10
Brazil	8511.46
Canada	8167.16
India	3463.36
Norway	2775.43
Japan	2723.67
Germany	2633.99

```
EDA (Exploratory Data Analysis)

# prompt: saya ingin tahu di kolom entity ada negara mana saja
unique_countries = df['Entity'].unique()
print(unique_countries)

['Afghanistan' 'Albania' 'Algeria' 'Angola' 'Antigua and Barbuda' 'Argentina' 'Armenia' 'Aruba' 'Australia' 'Austria' 'Azerbaijan' 'Bahamas' 'Bahrain' 'Bangladesh' 'Barbados' 'Belarus' 'Belgium' 'Belize' 'Benin' 'Bermuda' 'Bhutan' 'Bosnia and Herzegovina' 'Botswana' 'Brazil' 'Bulgaria' 'Burkina Faso' 'Burundi' 'Cambodia' 'Cameroon' 'Canada' 'Cayman Islands' 'Central African Republic' 'Chad' 'Chile' 'China' 'Colombia' 'Comoros' 'Congo' 'Costa Rica' 'Croatia' 'Cuba' 'Cyprus' 'Czechia' 'Denmark' 'Djibouti' 'Dominica' 'Dominican Republic' 'Ecuador' 'Egypt' 'El Salvador' 'Equatorial Guinea' 'Eritrea' 'Estonia' 'Eswatini' 'Ethiopia' 'Fiji' 'Finland' 'France' 'Gabon' 'Gambia' 'Georgia' 'Germany' 'Ghana' 'Greece' 'Grenada' 'Guatemala' 'Guinea' 'Guinea-Bissau' 'Guyana' 'Haiti' 'Hungary' 'Iceland' 'India' 'Indonesia' 'Iraq' 'Ireland' 'Israel' 'Italy' 'Jamaica' 'Japan' 'Jordan' 'Kazakhstan' 'Kenya' 'Kiribati' 'Kuwait' 'Kyrgyzstan' 'Latvia' 'Lebanon' 'Lesotho' 'Liberia' 'Libya' 'Lithuania' 'Luxembourg' 'Madagascar' 'Malawi' 'Malaysia' 'Maldives' 'Mali' 'Malta' 'Mauritania' 'Mauritius' 'Mexico' 'Mongolia' 'Montenegro' 'Morocco' 'Mozambique' 'Myanmar' 'Namibia' 'Nauru' 'Nepal' 'Netherlands' 'New Caledonia' 'New Zealand' 'Nicaragua' 'Niger' 'Nigeria' 'North Macedonia' 'Norway' 'Oman' 'Pakistan' 'Panama' 'Papua New Guinea' 'Paraguay' 'Peru' 'Philippines' 'Poland' 'Portugal' 'Puerto Rico' 'Qatar' 'Romania' 'Rwanda' 'Saint Kitts and Nevis' 'Saint Lucia' 'Saint Vincent and the Grenadines' 'Samoa' 'Sao Tome and Principe' 'Saudi Arabia' 'Senegal' 'Serbia' 'Seychelles' 'Sierra Leone' 'Singapore' 'Slovakia' 'Slovenia' 'Solomon Islands' 'Somalia' 'South Africa' 'South Sudan' 'Spain' 'Sri Lanka' 'Sudan' 'Suriname' 'Sweden' 'Switzerland' 'Tajikistan' 'Thailand' 'Togo' 'Tonga' 'Trinidad and Tobago' 'Tunisia' 'Turkey' 'Turkmenistan' 'Tuvalu' 'Uganda'
```

Data Pre-Processing

EDA

Modeling

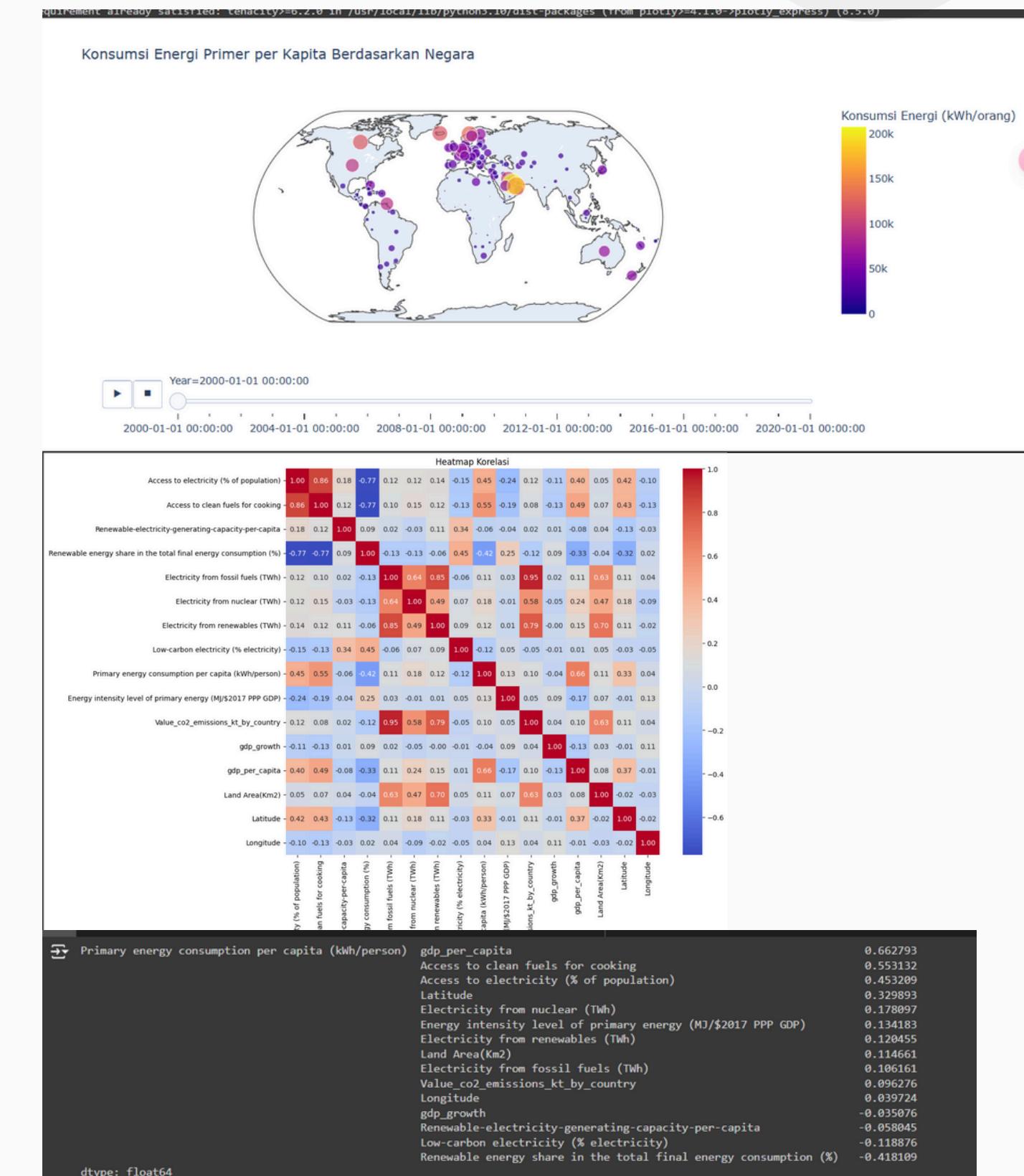
Recomendation

# Exploratory Data Analysis (EDA)

## "Analisa Insight di dalam data"

Terdapat beberapa poin yang berhasil dikulik setelah melakukan EDA:

1. Pemetaan penggunaan energy per KWH dari setiap negara menggunakan peta persebaran (menggunakan geopandas).  
Melalui geopandas, persebaran dari penggunaan energi dari yang terbesar hingga terkecil bisa terlihat dari warna dan besaran lingkaran. Selain itu kita bisa menambahkan filter untuk merubah tahun ke tahunnya. **Melalui peta ini dapat terlihat bahwa negara-negara timur tengah dan eropa merupakan negara dengan penggunaan energi terbesar.**
2. Melihat korelasi dari setiap kolom-kolom data menggunakan chart heatmaps (menggunakan seaborn). Berdasarkan chart heatmap tersebut dapat kesimpulan bahwa penggunaan energi paling tinggi itu berkorelasi dengan 3 hal :
  - gdp\_per\_capita
  - Access to clean fuels for cooking
  - Access to electricity (% of population)



Data Pre-Processing

EDA

Modeling

Recomendation

# Exploratory Data Analysis (EDA)

"Analisa Insight di dalam data"

## Insight

- Secara global, konsumsi energi primer per kapita cenderung meningkat. Visualisasi peta dunia menunjukkan peningkatan konsumsi di sebagian besar negara.
- Negara-negara berkembang seperti China dan India menunjukkan peningkatan konsumsi energi yang signifikan, mencerminkan pertumbuhan ekonomi dan industrialisasi mereka.

Data Pre-Processing

EDA

Modeling

Recomendation

## | Modeling

**Energy sources that are environmentally friendly, like renewables.**

---



# I Modeling

## "Multiple Time Series Forecasting"



Darts Modeling



Prophet Modeling



ARIMA Statistics Method

Data Pre-Processing

EDA

Modeling

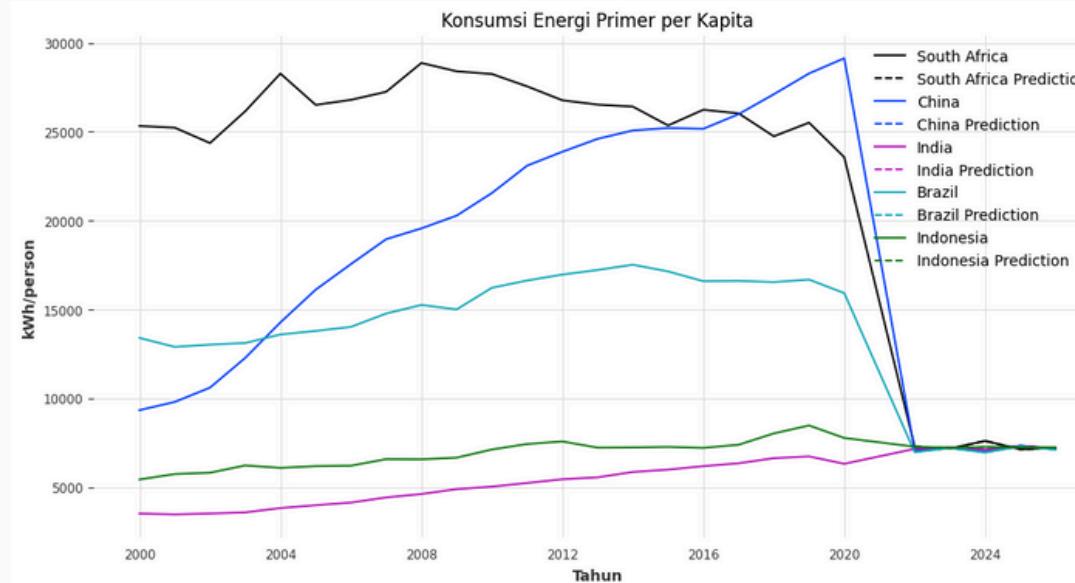
Recomendation

# I Modeling

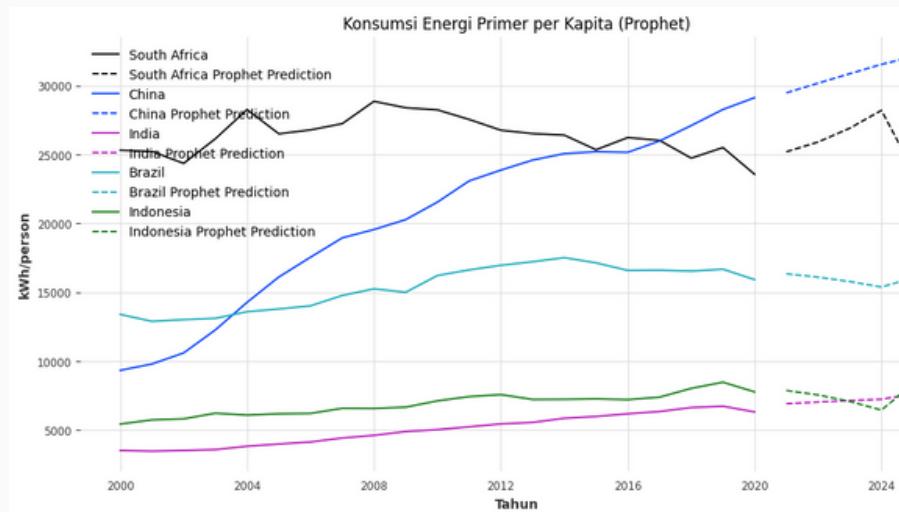
## "Multiple Time Series Forecasting"



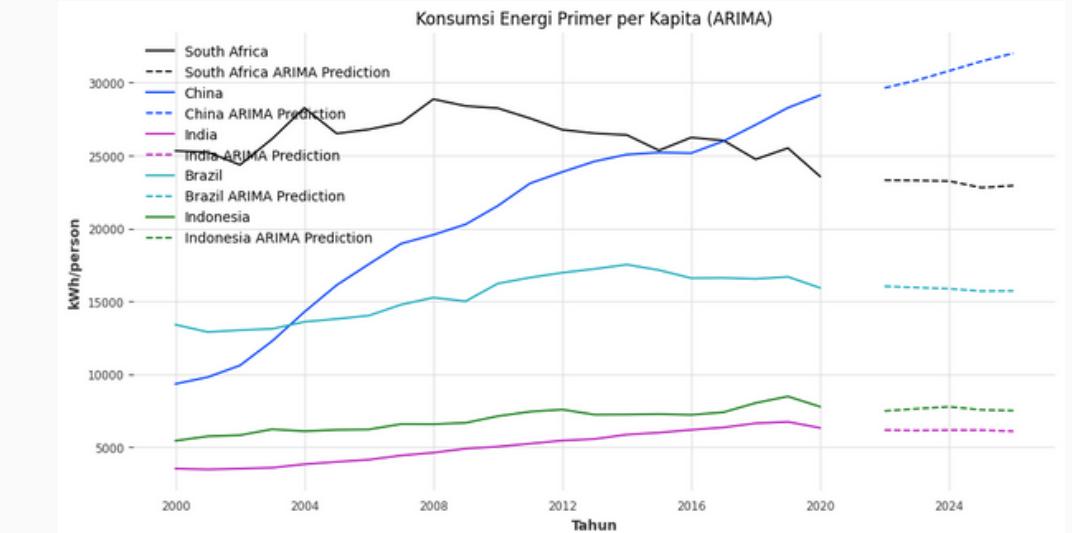
Darts Modeling



Prophet Modeling



ARIMA Statistics Method



## Keterangan

Mengambil 5 negara sebagai sampel, dengan mengambil negara dari setiap persebaran kumpulan energi terjauh: China, India, South Africa, Indonesia, dan Brazil. Memprediksi penggunaan energi untuk 5 tahun kedepan.

## Result Analisa

Secara grafik dapat terlihat bahwa prediksi yang cukup aneh terjadi pada pemodelan Darts, dengan drastis turun. Sedangkan pada pemodelan Prophet dan ARIMA masih normal dan dapat dimaklumi untuk prediksi time seriesnya

Data Pre-Processing

EDA

Modeling

Recomendation

# Modeling

## "Model Evaluation"



Darts Modeling

→ MAPE for South Africa: 71.10%  
MAPE for China: 73.51%  
MAPE for India: 11.83%  
MAPE for Brazil: 56.82%  
MAPE for Indonesia: 6.83%

Mean Absolute  
Percentage Error  
(MAPE)



Prophet Modeling

MAPE for South Africa (Prophet): 5.00%  
MAPE for China (Prophet): 13.80%  
MAPE for India (Prophet): 11.74%  
MAPE for Brazil (Prophet): 3.54%  
MAPE for Indonesia (Prophet): 10.92%



ARIMA Statistics Method

MAPE for South Africa (ARIMA): 8.21%  
MAPE for China (ARIMA): 13.72%  
MAPE for India (ARIMA): 4.55%  
MAPE for Brazil (ARIMA): 3.73%  
MAPE for Indonesia (ARIMA): 4.91%

## Keterangan

**MAPE** mengukur seberapa besar kesalahan peramalan relatif terhadap nilai sebenarnya, dalam bentuk persentase. Semakin kecil MAPE maka model tersebut semakin bagus/tepat

## Result Analisa

Melalui data ini dapat disimpulkan bahwa Model machine learning terbaik diantara ketiganya yaitu ARIMA Statistics Method, di susul dengan Prophet lalu ketiga ada Darts. Karena terlihat dari besaran MAPE nya yang paling kecil ada pada model ARIMA.

Data Pre-Processing

EDA

Modeling

Recomendation

## I Recommendation

**“A mindset focused on  
maintaining environmental  
balance for future generations.”**

---



# I Recommendation

## Recommendation

Berdasarkan analisis dan pemodelan yang dilakukan, berikut adalah beberapa rekomendasi:

- Negara-negara seperti China dan India menunjukkan peningkatan konsumsi energi primer per kapita. Investasi dalam sumber energi terbarukan di negara-negara ini dapat membantu memenuhi permintaan yang meningkat sambil mengurangi dampak lingkungan.
- Negara-negara dengan konsumsi energi primer per kapita yang tinggi dapat mengeksplorasi strategi efisiensi energi untuk mengurangi konsumsi secara keseluruhan.
- Pemodelan lebih lanjut dengan data yang lebih terperinci dan beragam dapat memberikan wawasan yang lebih akurat tentang tren konsumsi energi masa depan dan menginformasikan kebijakan energi.
- Kolaborasi internasional dalam penelitian dan pengembangan teknologi energi terbarukan sangat penting untuk transisi global menuju sistem energi yang berkelanjutan.

## Kesimpulan

Analisis data energi ini memberikan wawasan berharga tentang tren konsumsi energi global, sumber energi, dan korelasi antara faktor-faktor terkait. Rekomendasi yang disajikan di atas bertujuan untuk memandu pengambilan keputusan menuju masa depan energi yang lebih berkelanjutan dan adil. Implementasi rekomendasi ini memerlukan upaya bersama dari pemerintah, industri, dan masyarakat sipil.

Data Pre-Processing

EDA

Modeling

Recomendation

| Final Project Dibimbing.ID

# Thank You

Mari Jaga Bumi Kita



0851-7334-4907



arkhandoohan98@gmail.com

