

Laporan Data Analisis

PRODUKSI PADI DI PULAU SUMATERA (1993 – 2020)

Oleh :

Rendy Lutfi Prabowo

Juni 2025

Problem Definition

Padi merupakan komoditas strategis di Indonesia karena menjadi sumber pangan pokok bagi sebagian besar masyarakat. Pulau Sumatera termasuk salah satu lumbung padi nasional dengan kontribusi signifikan terhadap produksi beras nasional. Namun, produksi padi dapat dipengaruhi oleh berbagai faktor, mulai dari luas lahan, kondisi iklim (curah hujan, kelembapan, suhu), hingga kebijakan pertanian di masing-masing provinsi.

Analisis data produksi padi di Pulau Sumatera dari tahun ke tahun menjadi penting untuk :

- Memahami tren produksi dan fluktuasi yang terjadi.
- Mengidentifikasi provinsi penghasil utama serta wilayah dengan produktivitas tertinggi.
- Mengevaluasi sejauh mana faktor lingkungan berpengaruh terhadap hasil produksi.

Dengan analisis ini, dapat diperoleh insight yang bermanfaat untuk mendukung kebijakan pangan, meningkatkan efisiensi lahan, serta mengantisipasi penurunan produksi di masa depan.

Dataset Description

Dataset yang digunakan dalam analisis ini bersumber dari situs Kaggle (judul: [Dataset Tanaman Padi Sumatera](#)) yang diunggah oleh pengguna *ardikasatria*. Dataset mencakup data produksi padi di Pulau Sumatera dari tahun ke tahun, terdiri dari 224 baris dan 7 kolom. Kolom-kolom tersebut meliputi Provinsi, Tahun, Produksi (jumlah produksi padi), Luas Panen, Curah Hujan, Kelembapan, dan Suhu Rata-rata. Data ini memungkinkan analisis tren produksi, perbandingan antarprovinsi, serta hubungan antara faktor lingkungan dan hasil produksi.

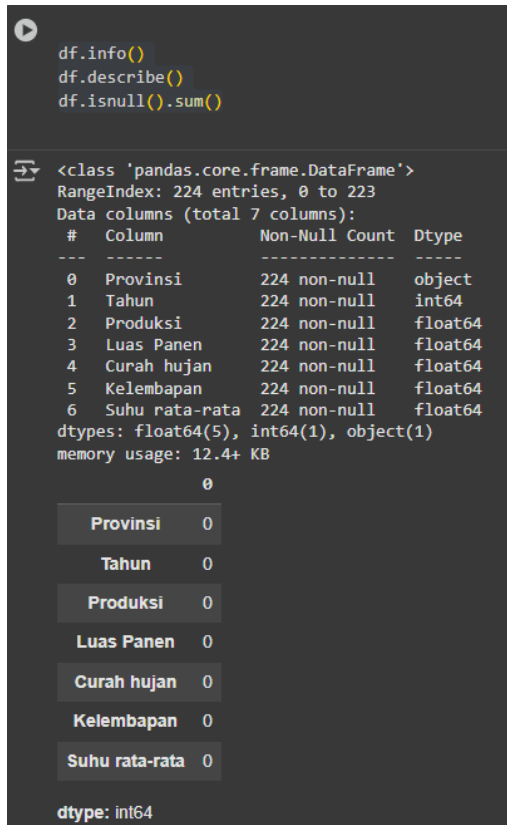
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load dataset
df = pd.read_csv('<sesuaikan dgn lokasi file csv di device anda>')
df.head()
```

Metodologi

- **Data Preparation** : cek struktur data, missing value, statistik deskriptif.

```
df.info()
df.describe()
df.isnull().sum()
```



```
df.info()
df.describe()
df.isnull().sum()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 224 entries, 0 to 223
Data columns (total 7 columns):
Column Non-Null Count Dtype
--- ---
0 Provinsi 224 non-null object
1 Tahun 224 non-null int64
2 Produksi 224 non-null float64
3 Luas Panen 224 non-null float64
4 Curah hujan 224 non-null float64
5 Kelembapan 224 non-null float64
6 Suhu rata-rata 224 non-null float64
dtypes: float64(5), int64(1), object(1)
memory usage: 12.4+ KB

Provinsi	0
Tahun	0
Produksi	0
Luas Panen	0
Curah hujan	0
Kelembapan	0
Suhu rata-rata	0

dtype: int64

- **Data Preprocessing** : encoding provinsi, normalisasi, perhitungan produktivitas.

```
# tidak ada data preprocessing di dataset ini, karena data sudah relatif rapi dan bersih
```

PRODUKSI PADI DI PULAU SUMATERA (1993 – 2020)

- **Exploratory Data Analysis (EDA)** : grafik tren, bar chart per provinsi, scatter luas panen vs produksi, heatmap korelasi.

Tren Produksi Padi per Tahun (EDA)

```
prod_per_tahun = df.groupby("Tahun")["Produksi"].sum()

plt.figure(figsize=(10,6))
sns.lineplot(x=prod_per_tahun.index, y=prod_per_tahun.values, marker="o")
plt.title("Tren Produksi Padi di Sumatera per Tahun")
plt.xlabel("Tahun")
plt.ylabel("Produksi")
plt.show()
```

Top Provinsi Penghasil Padi (EDA)

```
prod_per_prov = df.groupby("Provinsi")["Produksi"].mean().sort_values(ascending=False)

plt.figure(figsize=(12,6))
sns.barplot(x=prod_per_prov.values, y=prod_per_prov.index)
plt.title("Rata-rata Produksi Padi per Provinsi (1993-...)")
plt.xlabel("Produksi Rata-rata")
plt.ylabel("Provinsi")
plt.show()
```

Hubungan Produksi dan Luas Panen (EDA)

```
plt.figure(figsize=(8,6))
sns.scatterplot(x="Luas Panen", y="Produksi", hue="Provinsi", data=df)
plt.title("Hubungan Luas Panen vs Produksi")
plt.xlabel("Luas Panen")
plt.ylabel("Produksi")
plt.show()
```

Korelasi Faktor Lingkungan dengan Produksi (EDA)

```
plt.figure(figsize=(8,6))
sns.heatmap(df[["Produksi", "Luas Panen", "Curah hujan", "Kelembapan", "Suhu rata-rata"]].corr(),
            annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Korelasi Faktor Lingkungan dengan Produksi")
plt.show()
```

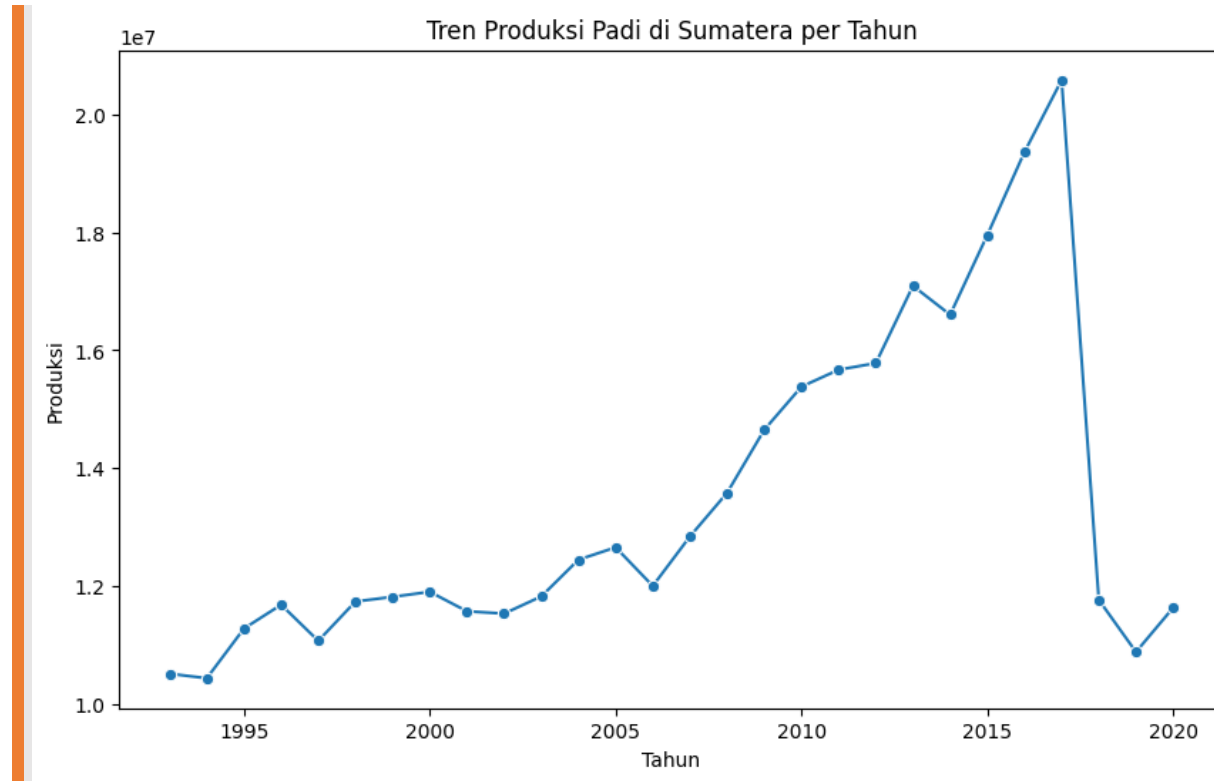
Produktivitas (Produksi per Luas Panen) (EDA)

```
df["Produktivitas"] = df["Produksi"] / df["Luas Panen"]
prod_per_prov = df.groupby("Provinsi")["Produktivitas"].mean().sort_values(ascending=False)

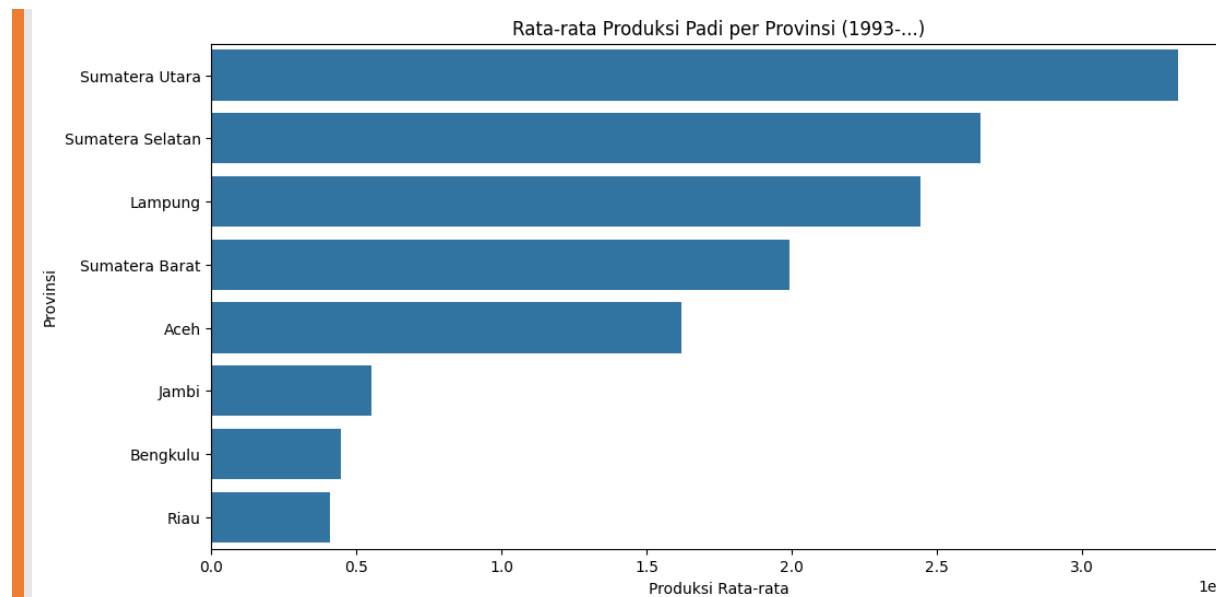
plt.figure(figsize=(12,6))
sns.barplot(x=prod_per_prov.values, y=prod_per_prov.index)
plt.title("Rata-rata Produktivitas Padi per Provinsi")
plt.xlabel("Produktivitas (Ton per Ha)")
plt.ylabel("Provinsi")
plt.show()
```

Hasil Analisis & Visualisasi

- **Tren produksi tahunan** (line chart).

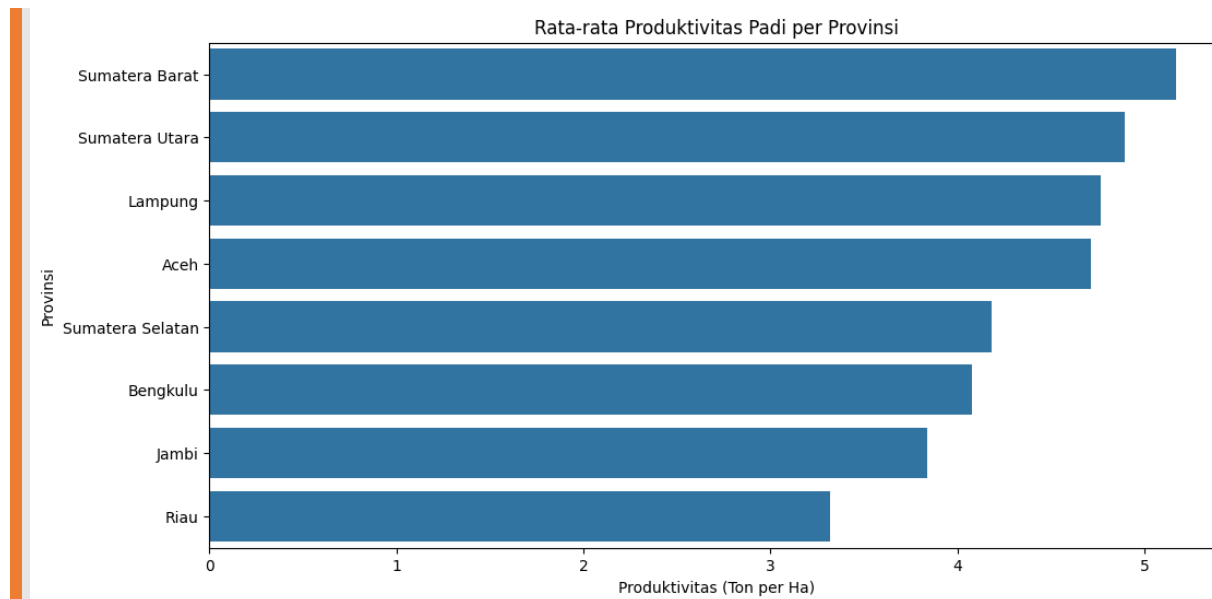


- **Provinsi penghasil terbesar** (bar chart).

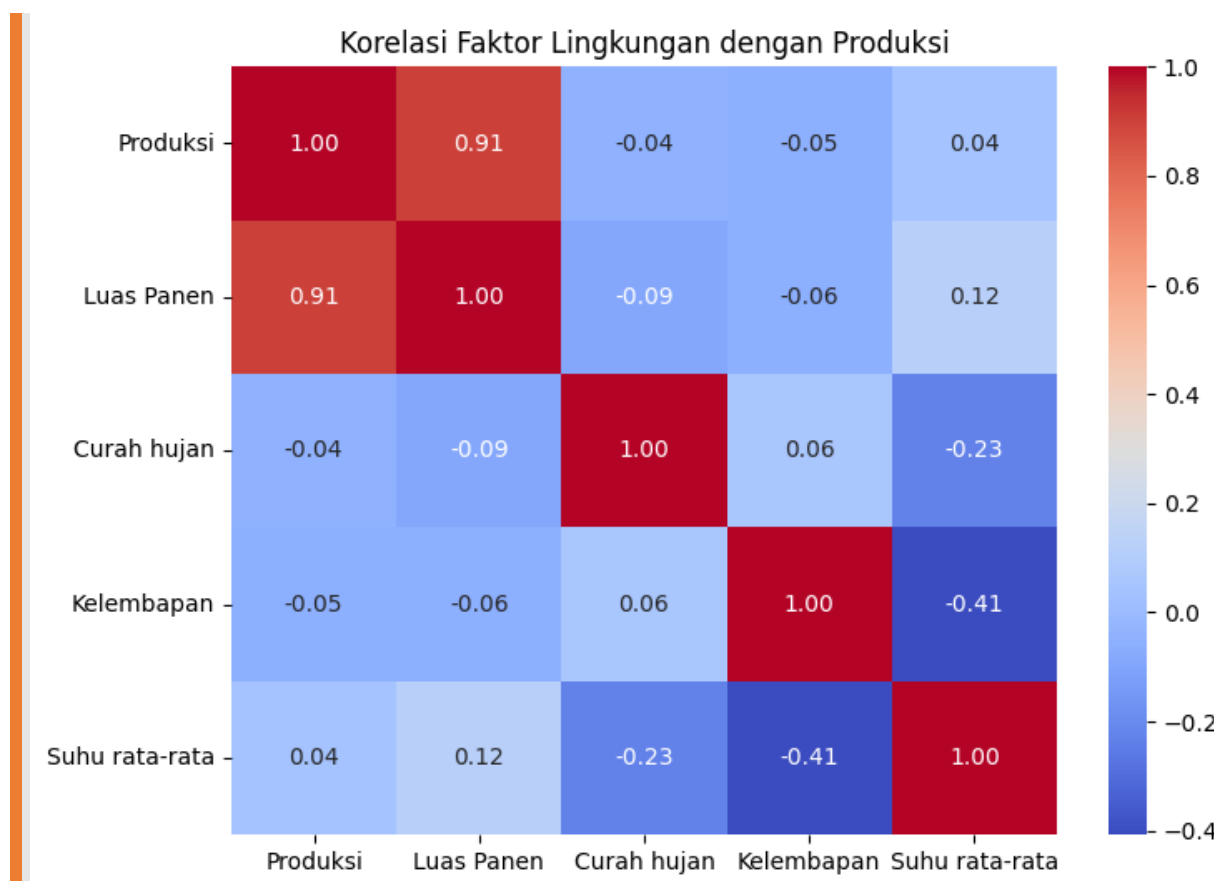


PRODUKSI PADI DI PULAU SUMATERA (1993 – 2020)

- **Produktivitas padi** (ton/ha) (bar chart).



- **Korelasi faktor lingkungan** (heatmap).



Kesimpulan

- Produksi padi di Sumatera tumbuh hingga 2017, tetapi mengalami penurunan drastis setelahnya.
- Sumatera Utara, Sumatera Selatan, dan Lampung merupakan pusat produksi utama, namun Sumatera Barat unggul dalam produktivitas per hektar.
- Faktor luas panen adalah penentu utama hasil produksi, sedangkan variabel iklim relatif kurang berpengaruh dalam data ini.

Rekomendasi

Kebijakan pertanian sebaiknya fokus pada efisiensi lahan dan peningkatan produktivitas (contoh: teknologi pertanian, varietas unggul) daripada hanya memperluas lahan.