



**Aplikasi Analisis Penduduk**

**Remastered 2025**

**Augmentasi • Imputasi • EDA • Prediksi • Streamlit**

# Proyek & Badge

License MIT

Python 3.8+

Streamlit

Scikit-Learn



# Overview

Aplikasi ini melakukan:

1. **Augmentasi** data penduduk dari 20→200 baris
2. **Imputasi** nilai kosong (Umur, Pendapatan, Status)
3. **EDA** interaktif & visualisasi
4. **Prediksi ML**: Status Nikah & Pendapatan
5. **Web App** dengan Streamlit



# Struktur Folder

```
penduduk\_analisis\_app/  
├── data/  
│   ├── dataset\_penduduk.csv  
│   ├── generated\_data.csv  
│   └── cleaned\_data.csv  
├── model/  
│   ├── model\_status\_nikah.pkl  
│   ├── model\_pendapatan.pkl  
│   └── label\_encoders.pkl  
├── output/  
│   └── eda\_visualization.png  
├── app.py  
├── generate\_data.py  
├── clean\_data.py  
├── eda.py  
├── train\_model.py  
└── requirements.txt
```



## Dataset Awal (20 Baris)

ID	Umur	Pendidikan	Pendapatan	Status Nikah
1	25	SMA	4.000.000	Belum
2	28	D3	4.500.000	Kawin
3	35	S1	—	Kawin
...	...	...	...	...
20	23	SMA	—	Belum



# Augmentasi Data

- Tambah **180 baris sintetis** → total  $\geq 200$
- **Umur**: 17–80 tahun
- **Pendidikan**: SMA, D3, S1, S2
- **Pendapatan** mengikuti distribusi per jenjang
- Simpan → `data/generated_data.csv`

```
# generate_data.py (snippet)
from faker import Faker
# ...
```



# Imputasi Nilai Kosong

- **Median** untuk Umur & Pendapatan
- **Random Forest Classifier** untuk Status Nikah
- Hasil → `data/cleaned_data.csv`

```
# clean_data.py (snippet)
from sklearn.impute import SimpleImputer
# ...
```



# EDA & Visualisasi

- Distribusi Pendidikan
- Rata-rata Pendapatan per Pendidikan
- Boxplot Pendapatan
- Scatterplot Umur vs Pendapatan
- Histogram Umur, Pie Status Nikah
- Export → `output/eda_visualization.png`





# Pelatihan Model

- **RandomForestClassifier** → Status Nikah
  - Akurasi: ~90%
- **RandomForestRegressor** → Pendapatan
  - $R^2$  Score: ~0.85
- Save:
  - `model/model_status_nikah.pkl`
  - `model/model_pendapatan.pkl`

## **Prediksi Individual**

1. Input: Umur & Pendidikan
2. Klik “Predict”
3. Tampil:
  - Status Nikah
  - Estimasi Pendapatan



# Prediksi Massal

1. Upload CSV (kolom: Umur, Pendidikan)
2. Model → massal
3. Download `hasil_prediksi_massal.csv`



# Tampilan Aplikasi



## Aplikasi Analisis & Prediksi Data Penduduk



### Statistik Deskriptif dan Visualisasi



### Statistik Numerik

	ID	Umur	Pendapatan
count	200	200	
mean	100.5	47.285	
std	57.8792	17.1758	
min	1	17	
25%	50.75	33.75	
50%	100.5	45	

# Teknologi

- **Python 3.8+**
- **Pandas, NumPy**
- **Scikit-learn**
- **Matplotlib, Seaborn, Plotly**
- **Streamlit**
- **Joblib / Pickle**



## Rekomendasi Fitur Tambahan

- Tambah kolom: jenis kelamin, pekerjaan
- REST API (FastAPI)
- Deployment Streamlit Cloud
- Integrasi DB (PostgreSQL / MongoDB)
- Versi mobile/app



## **Manfaat & Aplikasi**

- Studi kasus akhir semester
- Sensus & demografi
- Prototipe portofolio
- Penelitian sosial & ekonomi

# Terima Kasih & Kontak

“ Ingin lihat kode lengkap atau berdiskusi? ”

 GitHub [LycusCoder](#)

 LinkedIn [Nourivex](#)

 Email [affif@nourivex.tech](mailto:affif@nourivex.tech)



