

LAPORAN KOMPUTASI NUMERIK

Tugas Pemrograman A



Anggota Kelompok 24 :

Aliya Rizqiningrum Salamun	2306161813
Vanesa Kayla Zahra	2306161901
Calvin Wirathama Katoroy	2306242395

**FAKULTAS TEKNIK
PROGRAM STUDI TEKNIK KOMPUTER
UNIVERSITAS INDONESIA
DEPOK
MEI 2025**

SOAL NOMOR 1

1. Perkiraan nilai yang hilang

Untuk memperkirakan nilai yang hilang pada tahun 2005, 2006, 2015 dan 2016, (poin a-d) kami menggunakan metode interpolasi polinomial Newton. Metode ini dipilih karena kemampuannya untuk menangkap pertumbuhan non-linear dalam data penduduk dan pengguna internet Indonesia.

Perkiraan Jumlah Penduduk Indonesia

- Tahun 2005 & 2006

1. **Ekstraksi Data:** Menggunakan data dari tahun terdekat (2003, 2004, 2007, 2008)
2. **Metode:** Interpolasi Newton orde 3 untuk menangkap tren non-linear
3. **Hasil:**
 - Tahun 2005: **232.058.282 jiwa**
 - Tahun 2006: **235.157.743 jiwa**

- Tahun 2015 & 2016

1. **Ekstraksi Data:** Menggunakan data dari tahun terdekat (2013, 2014, 2017, 2018)
2. **Metode:** Interpolasi Newton orde 3
3. **Hasil:**
 - Tahun 2015: **261.333.583 jiwa**
 - Tahun 2016: **264.017.996 jiwa**

Perkiraan Persentase Pengguna Internet Indonesia

Untuk persentase pengguna internet (poin e-h), digunakan pendekatan regresi polinomial:

1. **Pemilihan Model:** Diuji berbagai derajat polinomial (2-6) untuk menemukan model optimal
2. **Evaluasi Model:** Menggunakan metrik R^2 , RMSE, MAE, dan analisis visual residual
3. **Implementasi:** Diprogram dengan Python di Jupyter Notebook

Hasil Akhir

Hasil Estimasi Nilai yang Hilang:

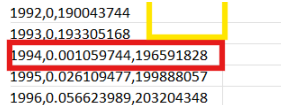
	Year	Estimated_Population	Estimated_Internet_Percentage
0	2005	2.320583e+08	2.637521
1	2006	2.351577e+08	3.359146
2	2015	2.613336e+08	24.680342
3	2016	2.640180e+08	29.599407

SOAL NOMOR 2

1. Persentase pengguna Internet Indonesia.

a) Langkah 1 : Identifikasi

Mencari persentase dan menggambarkan data visualisasinya membagi data menjadi 2 kelompok yaitu data populasi dan data internet. Data populasi dari tahun 1960 - 2023 dan data internet tahun 1994 - 2023 saat ada nilainya yang tidak null.



1992,0.190043744	
1993,0.193305168	
1994,0.001059744,196591828	
1995,0.026109477,199888057	
1996,0.056623989,203204348	

b) Langkah 2 : Memilih metode

Persamaan polinomial dan pencocokan kurva dapat melihat berbagai jenis pola dalam data yang memiliki waktu. Menentukan derajat untuk populasi yaitu 3 untuk dapat menangkap pola dengan baik. Sedangkan, persentase internet memiliki pola S-curve dengan derajat lebih tinggi yaitu 5. Metode yang digunakan yaitu regresi.

c) Langkah 3 : Mendetailkan algoritma dan pemrograman

Implementasi regresi polinomial untuk data excel yang ada lebih cocok diuji dengan beberapa derajat yaitu 2 - 5 untuk populasi dan 3 - 6 untuk internet. Kemudian, menggunakan fungsi `numpy.polyfit()` dan memilih model dengan RMSE terendah sesuai rumus. Validasi yang dilakukan untuk setiap model dan membandingkan nilai prediksi dengan nilai aktual pada beberapa sampel.

d) Langkah 4 : Membuat kode program

Penjelasan dan dokumentasi lengkap pada github.

e) Langkah 5 : Run Program

Menjalankan kode program pada jupyter notebook dan menampilkan nilai serta hasil visualisasi plot sebagai berikut :

```
Parameter model logistik: L=100.00, k=0.2493, x0=2019.48
Persamaan logistik persentase pengguna internet:
y = 100.00 / (1 + exp(-0.249287 * (x - 2019.48)))
R² model logistik untuk internet: 0.9971
```

2. Pertumbuhan populasi Indonesia.

a) Langkah 1 : Identifikasi

Seperti penjelasan diatas dengan langkah yang sama yaitu menggunakan metode regresi dan pencocokan polinomial dengan menentukan derajat yaitu 3.

b) Langkah 2 : Matematis

Menggunakan `numpy.polyfit` untuk mencari koefisien polinomial dengan meminimalkan jumlah kuadrat error antara model dengan data aktual. Dengan rumus :

$$y = a_0 + a_1X + a_2X^2 + \dots + a_nX^n$$

c) Langkah 3 : Run Program

Menjalankan kode program pada jupyter notebook dan menampilkan nilai serta hasil visualisasi plot sebagai berikut :

```
Derajat 2: R² = 0.9992291899369616
Derajat 3: R² = 0.999878721246272
Derajat 4: R² = 0.9998763139227992
Derajat 5: R² = 0.9998738888222108
Derajat polinomial terbaik untuk populasi: 3 dengan R² = 0.999878721246272
Persamaan polinomial populasi:
y = 2346982050909.58 - 3549267692.848241x + 1787567.680112x² - 299.812658x³
```

SOAL NOMOR 3

Hasil Estimasi

- Jumlah populasi Indonesia di tahun 2030: **295.384.447 jiwa**
- Jumlah pengguna Internet di Indonesia di tahun 2035: **297.010.262 jiwa**

Metodologi dan Langkah-Langkah Penyelesaian

1. Persiapan Data dan Tahun Prediksi
 - Membuat array tahun prediksi **[2030, 2035]** dan mereformatnya menjadi array 2D
 - Menggunakan **reshape(-1, 1)** untuk menyesuaikan format data yang dibutuhkan model sklearn
2. Pemodelan dan Prediksi Populasi
 - Menerapkan transformasi polinomial pada tahun prediksi dengan **poly_features_pop.transform(future_years)**
 - Melakukan prediksi populasi dengan model polinomial yang telah dilatih: **pop_model.predict()**
 - Model populasi menggunakan persamaan:
$$y = 2346982050909.58 - 3549267692.848241x + 1787567.680112x^2 - 299.812658x^3$$
 - Model memiliki tingkat kecocokan sangat tinggi dengan data historis ($R^2 = 0.9998$)
3. Pemodelan dan Prediksi Pengguna Internet
 - Menggunakan model logistik untuk memprediksi persentase pengguna internet
 - Prediksi menunjukkan **97,96%** populasi Indonesia akan menjadi pengguna internet pada 2035
4. Perhitungan Jumlah Absolut Pengguna Internet
 - Mengkonversi persentase menjadi jumlah absolut dengan rumus:
Jumlah pengguna = (Persentase/100) × Populasi
 - Menghasilkan prediksi **297.010.262** pengguna internet di tahun 2035
5. Visualisasi Hasil
 - Membuat visualisasi untuk model pertumbuhan populasi : Menampilkan data aktual, kurva model polinomial, dan titik prediksi
 - Membuat visualisasi untuk model pertumbuhan persentase pengguna internet : Menampilkan data aktual, kurva model logistik, dan titik prediksi
6. Analisis dan Interpretasi
 - Model polinomial berhasil menangkap tren pertumbuhan populasi dengan akurasi tinggi
 - Model logistik untuk penetrasi internet menunjukkan tren saturasi mendekati 100% pada 2035

- Hasil prediksi disusun dalam DataFrame untuk memudahkan interpretasi dan pelaporan

Prompt yang dipakai

K Halo, saya sedang mencoba mengerjakan tugas pemodelan data untuk memprediksi populasi Indonesia dan persentase pengguna internet. Saya sudah mencoba menggunakan regresi linear sederhana, tapi hasilnya tidak terlalu akurat. Data internet terlihat seperti kurva S, dan saya bingung metode apa yang paling tepat untuk modelnya. Bisakah Anda memberikan beberapa saran tentang pendekatan yang sebaiknya saya gunakan? Saya khususnya bingung tentang:

1. Bagaimana cara memilih model yang tepat untuk data populasi vs data persentase internet?
2. Apakah sebaiknya saya gunakan teknik yang berbeda untuk kedua jenis data?
3. Bagaimana cara menilai keakuratan model yang saya buat?

Evaluating Model Accuracy

1. **Statistical metrics:**
 - R² for overall fit quality
 - RMSE/MAE for error magnitude
 - AIC/BIC for model comparison while penalizing complexity
2. **Visual validation:**
 - Plot predictions against actual data
 - Check residuals for patterns/heteroscedasticity
 - Pay special attention to recent trends - do they match your model?
3. **Cross-validation:**
 - Time-series cross-validation (not random CV)
 - Predict known values by training on subsets of historical data

For the internet usage model, ensure predictions stay within logical bounds (0-100%) and follow expected adoption patterns.

Guidance on Modeling Population and Internet Usage Data

Selecting Appropriate Models

For Population Data:

- Population growth typically follows more regular patterns than technology adoption
- Exponential regression might be appropriate after ensuring data follows that growth pattern
- Linear models might work for short-term extrapolation but miss long-term trends
- Logistic or S-curve models can be appropriate if you expect population growth to eventually plateau

For Internet Usage Data:

- Logistic regression is ideal for S-shaped adoption curves with natural upper bounds (100%)
- Consider segmenting by region (e.g., urban vs rural) as adoption rates differ
- Pay attention to recent trends (2010-2023) as they may deviate from historical patterns
- An S-curve model is often the best fit for percentage data
- 1-growth rate
- An S-curve model is often the best fit for percentage data

Using Different Techniques for Each Dataset

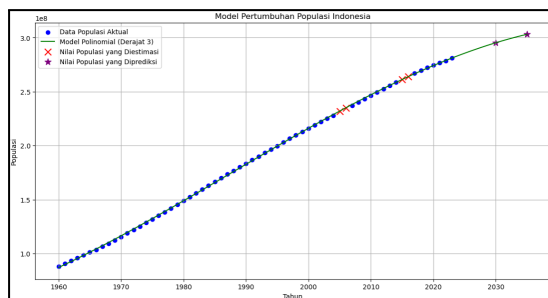
The goal is to tailor your approach to each dataset's characteristics:

- Population data: More traditional statistical models (polynomial regression, exponential smoothing)
- Internet percentages: S-curve/Logistic models that handle bounded variables much more effectively

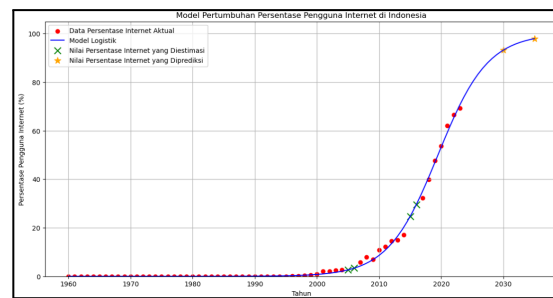
VISUALISASI

Visualisasi Data dan Plot

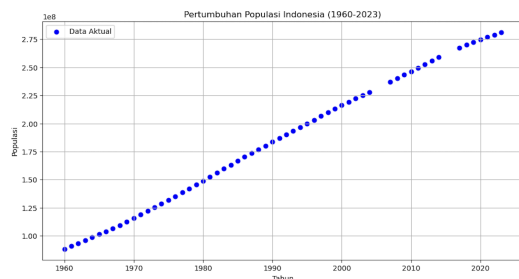
Visualisasi Model Pertumbuhan Populasi



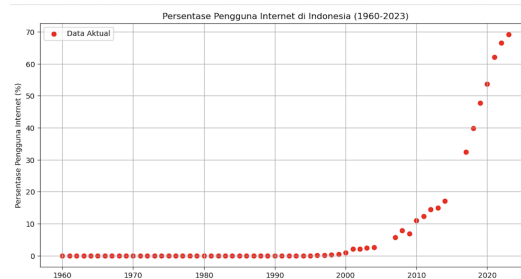
Visualisasi Model Persentase Pengguna Internet



Visualisasi Model Persentase Pengguna Internet



Visualisasi Persentase Pengguna Internet di Indonesia (1960 - 2023)



Link Github Repository

Link Github Repository: <https://github.com/aliyars/komnum-tugas-pemroga>