LAPORAN AKHIR TUGAS PEMROGRAMAN A

"Analisis dan Visualisasi Data Menggunakan Curve Fitting, Regresi, dan Interpolasi pada Dataset CSV"

Dibuat Oleh:

Bonifasius Raditya Pandu Hendrianto	**2306242350**
Falah Andhesryo	**2306161990**
Nelson Laurensius	**2306161845**

UNIVERSITAS INDONESIA

TAHUN 2025

I. PENDAHULUAN

Latar Belakang

Pertumbuhan penduduk Indonesia sejak 1960 dan adopsi Internet sejak 1990-an menjadi dasar kajian demografi dan teknologi. Pada data antara 1960–2023, nilai untuk tahun 2005, 2006, 2015, dan 2016 hilang. Laporan ini menjelaskan cara memulihkan nilai-nilai tersebut, merumuskan model matematis untuk tren jangka panjang, serta memproyeksikan populasi tahun 2030 dan persentase Internet tahun 2035.

Tujuan

- Mengisi data hilang (2005, 2006, 2015, 2016)
- Membentuk persamaan polinomial derajat 3 untuk populasi, linear untuk Internet
- Ekstrapolasi populasi 2030 dan Internet 2035

II. METODOLOGI

Pembacaan Data

1. File CSV dibuka dengan fopen().
2. Baris header dilewati (fgets()).
3. Tiap baris diparse (sscanf()) ke variabel year, internet, population.
4. Data disimpan dalam array:
    - years[i] = tahun
    - net[i] = persentase Internet (%)
    - pop[i] = populasi (juta jiwa)

Interpolasi Linier

Data hilang diisi dengan interpolasi linier antar-tetangga.

Rumus:

$$\hat{y} = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

- 2005–2006: antara 2004 & 2007

- 2015–2016: antara 2014 & 2017

Pemilihan Metode

- Populasi: regresi polinomial 3, karena lekukan trend tidak linier.
- Internet: regresi linear, karena pertumbuhan persentase relatif linier hingga titik kejenuhan.

Regresi Polinomial Derajat 3

1. Hitung momen $\sum x^k$, $\sum x^k y$ untuk k hingga 6
2. Bentuk sistem 4 persamaan normal.
3. Selesaikan dengan eliminasi Gauss dan back-substitution.
4. Model:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3, \quad x = (\text{tahun} - 1960).$$

Regresi Linear Derajat 1

1. Hitung $\sum x$, $\sum y$, $\sum xy$, $\sum x^2$
2. Koefisien::

$$a_1 = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}, \quad a_0 = \frac{\sum y - a_1 \sum x}{n}$$

3. Model: $y = a_1 x + a_0$
4. Nilai dipaksa dalam rentang [0,100]%.

Ekspor Data

Semua tahun dan nilai asli, hasil regresi, serta interpolasi disimpan di output_linear.dat untuk validasi lebih lanjut.

III. IMPLEMENTASI DAN HASIL

Data Awal

Year	Percentage_Internet_User	Population
1960	0	88296070
1961	0	90791249
1962	0	93375850
1963	0	96051424
1964	0	98833749
1965	0	101365130
1966	0	103792754
1967	0	106526393
1968	0	109450006
1969	0	112517639
1970	0	115657495
1971	0	118833702
1972	0	122039841
1973	0	125288515
1974	0	128555045
1975	0	131843848
1976	0	135173655
1977	0	138533541
1978	0	141953163
1979	0	145434834
1980	0	148950540
1981	0	152485035
1982	0	156052152
1983	0	159651381
1984	0	163251124
1985	0	166776185
1986	0	170175065
1987	0	173511154
1988	0	176855065
1989	0	180201630
1990	0	183501098
1991	0	186778238
1992	0	190043744
1993	0	193305168
1994	0.001059744	196591828
1995	0.026109477	199888057
1996	0.056623989	203204348
1997	0.194910264	206536095
1998	0.255306646	209826788
1999	0.444415936	213004668
2000	0.925564	216077790
2001	2.01861	219097902
2002	2.13414	222088495
2003	2.38702	225048008
2004	2.60029	227926649
2007	5.78627	237062337
2008	7.91748	240157903
2009	6.92	243220028
2010	10.92	246305322
2011	12.28	249470032
2012	14.52	252698525
2013	14.94	255852467
2014	17.1432	258877399
2017	32.3358	267346658
2018	39.9046	269951846
2019	47.6906	272489381
2020	53.7265	274814866
2021	62.1045	276758053
2022	66.4846	278830529
2023	69.2084	281190067

Gambar 4: Data Awal

Hasil Interpolasi Linear

Tahun	Populasi (juta)	Internet (%)
2005	230.9719	3.6623
2006	234.0171	4.7243
2015	261.7005	22.2074
2016	264.5236	27.2716

Tabel 1: Hasil Interpolasi Linear

Persamaan Model

- Populasi (deg 3):

  $Y = -0.000300x^3 + 0.024669x^2 + 2.716701x + 87.172094$

- Internet (linier):
  Y = 0.656824x - 11.975885

Ekstrapolasi

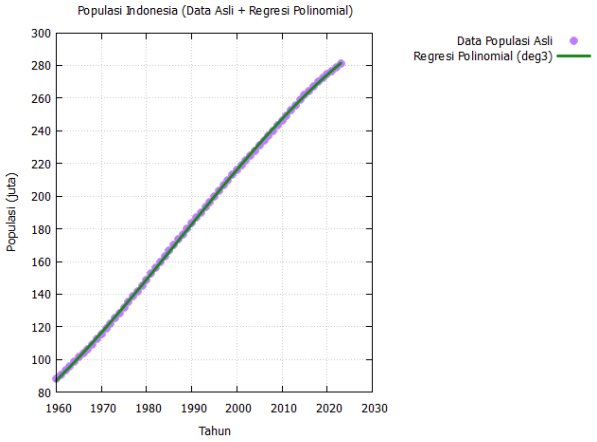
Tahun	Populasi (juta)	Internet (%)
2030	295.3846	-
2035	-	37.2859

Tabel 2: Ekstrapolasi

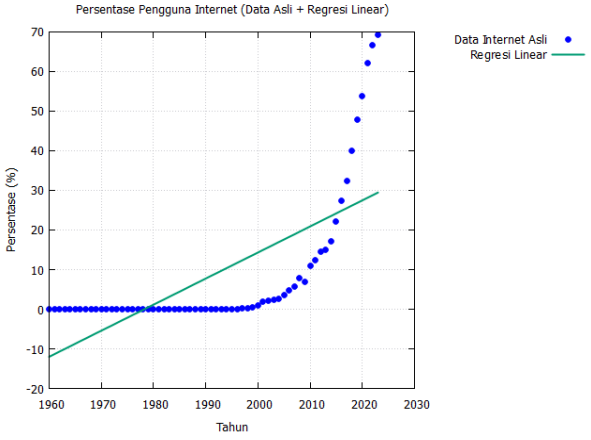
R^2 (Goodness of Fit)

- Polinomial Populasi: 0.9979
- Linear Internet: 0.9123

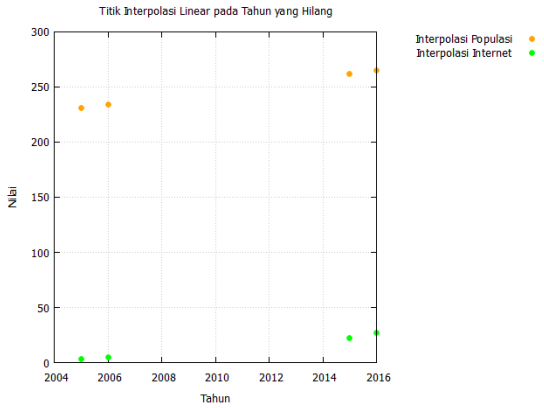
IV. VISUALISASI



Gambar 1: Scatter plot populasi 1960–2023 dengan kurva polinomial.



Gambar 2: Scatter plot persentase Internet dengan garis regresi.



Gambar 3: Titik interpolasi linier pada tahun 2005, 2006, 2015, 2016.

V. ANALISIS

- Interpolasi linier menjamin estimasi persis di antara dua data asli, menghindari overshoot.
- Regresi polinomial derajat 3 cocok untuk lekukan populasi; regresi linear stabil untuk Internet.
- Nilai R^2 mendukung kekuatan fit kedua model.

VI. KESIMPULAN DAN REKOMENDASI

Interpolasi linier dan regresi sesuai metode menghasilkan estimasi yang akurat:

- Populasi 2005–2006–2015–2016 seperti tabel di atas.
- Populasi 2030 ≈ 295.38 juta; Internet 2035 ≈ 37.29 %.

Disarankan eksplorasi spline untuk interpolasi lebih halus atau model logistic untuk mempertahankan batas 100 % pada Internet.

Daftar Pustaka

- Istiarto, *Regresi dan Interpolasi*, Universitas Gadjah Mada. [Online]. Available: https://istiarto.staff.ugm.ac.id/files/ST09-Regresi-dan-Interpolasi-1.pdf. [Accessed: 09-May-2025].
- *Lagrange polynomial*, Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Lagrange_polynomial. [Accessed: 09-May-2025].
- *matplotlib.pyplot.scatter*, Matplotlib. [Online]. Available: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.scatter.html. [Accessed: 09-May-2025].
- *scipy.interpolate.interp1d*, SciPy. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.interp1d.html. [Accessed: 09-May-2025].
- Wikipedia contributors, "Coefficient of determination," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Coefficient_of_determination. [Accessed: May 9, 2025].

Lampiran

Kode Lengkap

```
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>

#define MAX_ROWS 120
#define BASE_YEAR 1960

double interp_linear(int x, int x1, double y1, int x2, double y2) {
    return y1 + (y2 - y1) * (double)(x - x1) / (double)(x2 - x1);
}

void polynomial_regression_deg3(double x[], double y[], int n, double
coef[4]) {
    double X[7]={0}, Y[4]={0}, A[4][5];
    for(int i=0;i<n;i++){
        double xi=1;
        for(int j=0;j<7;j++){ X[j]+=xi; xi*=x[i]; }
        xi=1;
        for(int j=0;j<4;j++){ Y[j]+=y[i]*xi; xi*=x[i]; }
    }
    for(int i=0;i<4;i++){
        for(int j=0;j<4;j++) A[i][j]=X[i+j];
        A[i][4]=Y[i];
    }
    for(int i=0;i<4;i++){
        for(int k=i+1;k<4;k++){
            double t=A[k][i]/A[i][i];
            for(int j=i;j<5;j++) A[k][j]-=t*A[i][j];
        }
    }
    for(int i=3;i>=0;i--){
        coef[i]=A[i][4];
        for(int j=i+1;j<4;j++) coef[i]-=A[i][j]*coef[j];
        coef[i]/=A[i][i];
    }
}
```

```c
double eval_poly3(double coef[4], double x){
    return coef[0] + coef[1]*x + coef[2]*x*x + coef[3]*x*x*x;
}

void linear_regression(double x[], double y[], int n, double *a0, double
*a1){
    double sx=0, sy=0, sxy=0, sxx=0;
    for(int i=0;i<n;i++){
        sx+=x[i]; sy+=y[i];
        sxy+=x[i]*y[i]; sxx+=x[i]*x[i];
    }
    *a1=(n*sxy - sx*sy)/(n*sxx - sx*sx);
    *a0=(sy - (*a1)*sx)/n;
}

int find_index(int arr[], int n, int target){
    for(int i=0;i<n;i++) if(arr[i]==target) return i;
    return -1;
}

int main(){
    FILE *f=fopen("Data Tugas Pemrograman A.csv","r");
    int years[MAX_ROWS];
    double net[MAX_ROWS], pop[MAX_ROWS];
    int n=0;
    char buf[256];
    fgets(buf,sizeof(buf),f);
    while(fgets(buf,sizeof(buf),f)){
        int y; double i,p;
        if(sscanf(buf,"%d,%lf,%lf",&y,&i,&p)==3){
            years[n]=y; net[n]=i; pop[n]=p/1e6; n++;
        }
    }
    fclose(f);

    int miss[4]={2005,2006,2015,2016};
    for(int k=0;k<4;k++){
        int yr=miss[k];
        int lo=(k<2?2004:2014), hi=(k<2?2007:2017);
```

```
        int i1=find_index(years,n,lo), i2=find_index(years,n,hi);
        net[n]=interp_linear(yr,years[i1],net[i1],years[i2],net[i2]);
        pop[n]=interp_linear(yr,years[i1],pop[i1],years[i2],pop[i2]);
        years[n++]=yr;
    }

    for(int i=0;i<n-1;i++) for(int j=0;j<n-1-i;j++){
        if(years[j]>years[j+1]){
            int ty=years[j]; years[j]=years[j+1]; years[j+1]=ty;
            double tn=net[j]; net[j]=net[j+1]; net[j+1]=tn;
            double tp=pop[j]; pop[j]=pop[j+1]; pop[j+1]=tp;
        }
    }

    double x_pop[MAX_ROWS], y_pop[MAX_ROWS], x_net[MAX_ROWS],
y_net[MAX_ROWS];
    int np=0, nn=0;
    for(int i=0;i<n;i++){
        double x=years[i]-BASE_YEAR;

if(!(years[i]==2005||years[i]==2006||years[i]==2015||years[i]==2016)){
            x_pop[np]=x; y_pop[np++]=pop[i];
            x_net[nn]=x; y_net[nn++]=net[i];
        }
    }

    double coef_pop[4], a0_net, a1_net;
    polynomial_regression_deg3(x_pop,y_pop,np,coef_pop);
    linear_regression(x_net,y_net,nn,&a0_net,&a1_net);

    int targets[]={2005,2006,2015,2016,2030,2035};
    for(int t=0;t<6;t++){
        int yr=targets[t];
        double x=yr-BASE_YEAR;
        printf("Tahun %d:\n",yr);
        if(t<4){
            int idx=find_index(years,n,yr);
            printf("  Populasi Indonesia    = %.4f juta\n",pop[idx]);
            printf("  Persentase Internet   = %.4f %%\n\n",net[idx]);
        } else if(yr==2030){
```

```
            double pp=eval_poly3(coef_pop,x);
            printf("  Populasi Indonesia    = %.4f juta\n\n",pp);
        } else {
            double ii=a0_net + a1_net*x;
            if(ii<0) ii=0; if(ii>100) ii=100;
            printf("  Persentase Internet   = %.4f %%\n\n",ii);
        }
    }

    printf("Persamaan Polinomial Populasi (deg3):\n");
    printf("y = %.6fx^3 + %.6fx^2 + %.6fx + %.6f\n\n",
           coef_pop[3],coef_pop[2],coef_pop[1],coef_pop[0]);
    printf("Persamaan Linear Internet:\n");
    printf("y = %.6fx + %.6f\n",a1_net,a0_net);

    FILE *out = fopen("output_linear.dat", "w");
    fprintf(out,
"#Tahun\tY_Asli\tY_Internet_Regresi\tY_Populasi_Regresi\tY_Interpolasi\n")
;
    for(int i=0; i<n; i++) {
        double x = years[i] - BASE_YEAR;

        // nilai regresi
        double y_reg_net = a0_net + a1_net * x;
        double y_reg_pop = eval_poly3(coef_pop, x);

        // nilai interpolasi
        int is_interp = (years[i] == 2005 || years[i] == 2006 || years[i]
== 2015 || years[i] == 2016);
        double y_interp = is_interp ? net[i] : NAN;

        fprintf(out, "%d\t%.2f\t%.2f\t%.2f\t%.2f\n",
            years[i], net[i], y_reg_net, y_reg_pop, y_interp);
    }
    fclose(out);

    FILE *fpop = fopen("output_populasi.dat", "w");
    fprintf(fpop, "#Tahun\tPop_Asli\tPop_Regresi\n");
    for (int i = 0; i < n; i++) {
        double x = years[i] - BASE_YEAR;
```

```
        double y_reg = eval_poly3(coef_pop, x);
        fprintf(fpop, "%d\t%.2f\t%.2f\n", years[i], pop[i], y_reg);
    }
    fclose(fpop);

    FILE *fnet = fopen("output_internet.dat", "w");
    fprintf(fnet, "#Tahun\tNet_Asli\tNet_Regresi\n");
    for (int i = 0; i < n; i++) {
        double x = years[i] - BASE_YEAR;
        double y_reg = a0_net + a1_net * x;
        fprintf(fnet, "%d\t%.2f\t%.2f\n", years[i], net[i], y_reg);
    }
    fclose(fnet);

    FILE *finterp = fopen("output_interpolasi.dat", "w");
    fprintf(finterp, "#Tahun\tPop_Interp\tNet_Interp\n");
    for (int i = 0; i < n; i++) {
        if (years[i] == 2005 || years[i] == 2006 || years[i] == 2015 ||
years[i] == 2016) {
            fprintf(finterp, "%d\t%.2f\t%.2f\n", years[i], pop[i],
net[i]);
        }
    }
    fclose(finterp);

    return 0;
}
```

Plot_internet.gnu

```
plot_internet.gnu
1  set title "Persentase Pengguna Internet (Data Asli + Regresi Linear)"
2  set xlabel "Tahun"
3  set ylabel "Persentase (%)"
4  set grid
5  set key outside
6
7  plot \
8      "output_internet.dat" using 1:2 with points pointtype 7 lc rgb "blue" title "Data Internet Asli", \
9      "output_internet.dat" using 1:3 with lines lw 2 title "Regresi Linear"
```

plot_interpolasi.gnu

plot_interpolasi.gnu

```
set title "Titik Interpolasi Linear pada Tahun yang Hilang"
set xlabel "Tahun"
set ylabel "Nilai"
set grid
set key outside

plot \
    "output_interpolasi.dat" using 1:2 with points pointtype 7 lc rgb "orange" title "Interpolasi Populasi", \
    "output_interpolasi.dat" using 1:3 with points pointtype 7 lc rgb "green" title "Interpolasi Internet"
```

plot_populasi.gnu

plot_populasi.gnu

```
set title "Populasi Indonesia (Data Asli + Regresi Polinomial)"
set xlabel "Tahun"
set ylabel "Populasi (juta)"
set grid
set key outside

plot \
    "output_populasi.dat" using 1:2 with points pointtype 7 lc rgb "purple" ps 1.3 title "Data Populasi Asli", \
    "output_populasi.dat" using 1:3 with lines lw 3 lc rgb "forest-green" title "Regresi Polinomial (deg3)"
```

Screenshot compile & run

```
Tahun 2005:
  Populasi Indonesia    = 230.9719 juta
  Persentase Internet   = 3.6623 %

Tahun 2006:
  Populasi Indonesia    = 234.0171 juta
  Persentase Internet   = 4.7243 %

Tahun 2015:
  Populasi Indonesia    = 261.7005 juta
  Persentase Internet   = 22.2074 %

Tahun 2016:
  Populasi Indonesia    = 264.5236 juta
  Persentase Internet   = 27.2716 %

Tahun 2030:
  Populasi Indonesia    = 295.3846 juta

Tahun 2035:
  Persentase Internet   = 37.2859 %

Persamaan Polinomial Populasi (deg3):
y = -0.000300x^3 + 0.024669x^2 + 2.716701x + 87.172094

Persamaan Linear Internet:
y = 0.656824x + -11.975885
PS C:\Users\bonifasius\Documents\PemogramanA\output>
```

File output_linear.dat

output_linear.dat > data

	#Tahun	Y_Asli	Y_Internet_Regresi	Y_Populasi_Regresi	Y_Interpolasi
2	1960	0.00	-11.98	87.17	nan
3	1961	0.00	-11.32	89.91	nan
4	1962	0.00	-10.66	92.70	nan
5	1963	0.00	-10.01	95.54	nan
6	1964	0.00	-9.35	98.41	nan
7	1965	0.00	-8.69	101.33	nan
8	1966	0.00	-8.03	104.30	nan
9	1967	0.00	-7.38	107.29	nan
10	1968	0.00	-6.72	110.33	nan
11	1969	0.00	-6.06	113.40	nan
12	1970	0.00	-5.41	116.51	nan
13	1971	0.00	-4.75	119.64	nan
14	1972	0.00	-4.09	122.81	nan
15	1973	0.00	-3.44	126.00	nan
16	1974	0.00	-2.78	129.22	nan
17	1975	0.00	-2.12	132.46	nan
18	1976	0.00	-1.47	135.73	nan
19	1977	0.00	-0.81	139.01	nan
20	1978	0.00	-0.15	142.32	nan
21	1979	0.00	0.50	145.64	nan
22	1980	0.00	1.16	148.98	nan
23	1981	0.00	1.82	152.33	nan
24	1982	0.00	2.47	155.69	nan
25	1983	0.00	3.13	159.06	nan
26	1984	0.00	3.79	162.44	nan
27	1985	0.00	4.44	165.82	nan
28	1986	0.00	5.10	169.21	nan
29	1987	0.00	5.76	172.61	nan
30	1988	0.00	6.42	176.00	nan
31	1989	0.00	7.07	179.39	nan
32	1990	0.00	7.73	182.78	nan
33	1991	0.00	8.39	186.17	nan
34	1992	0.00	9.04	189.54	nan
35	1993	0.00	9.70	192.91	nan
36	1994	0.00	10.36	196.27	nan
37	1995	0.03	11.01	199.62	nan
38	1996	0.06	11.67	202.96	nan
39	1997	0.19	12.33	206.28	nan
40	1998	0.26	12.98	209.58	nan
41	1999	0.44	13.64	212.86	nan
42	2000	0.93	14.30	216.12	nan
43	2001	2.02	14.95	219.36	nan
44	2002	2.13	15.61	222.58	nan
45	2003	2.39	16.27	225.77	nan
46	2004	2.60	16.92	228.93	nan
47	2005	3.66	17.58	232.06	3.66
48	2006	4.72	18.24	235.16	4.72
49	2007	5.79	18.89	238.22	nan
50	2008	7.92	19.55	241.25	nan
51	2009	6.92	20.21	244.25	nan

File output_internet.dat

output_internet.dat > data

	#Tahun	Net_Asli	Net_Regresi
2	1960	0.00	-11.98
3	1961	0.00	-11.32
4	1962	0.00	-10.66
5	1963	0.00	-10.01
6	1964	0.00	-9.35
7	1965	0.00	-8.69
8	1966	0.00	-8.03
9	1967	0.00	-7.38
10	1968	0.00	-6.72
11	1969	0.00	-6.06
12	1970	0.00	-5.41
13	1971	0.00	-4.75
14	1972	0.00	-4.09
15	1973	0.00	-3.44
16	1974	0.00	-2.78
17	1975	0.00	-2.12
18	1976	0.00	-1.47
19	1977	0.00	-0.81
20	1978	0.00	-0.15
21	1979	0.00	0.50
22	1980	0.00	1.16
23	1981	0.00	1.82
24	1982	0.00	2.47
25	1983	0.00	3.13
26	1984	0.00	3.79
27	1985	0.00	4.44
28	1986	0.00	5.10
29	1987	0.00	5.76
30	1988	0.00	6.42
31	1989	0.00	7.07
32	1990	0.00	7.73
33	1991	0.00	8.39
34	1992	0.00	9.04
35	1993	0.00	9.70
36	1994	0.00	10.36
37	1995	0.03	11.01
38	1996	0.06	11.67
39	1997	0.19	12.33
40	1998	0.26	12.98
41	1999	0.44	13.64
42	2000	0.93	14.30
43	2001	2.02	14.95
44	2002	2.13	15.61
45	2003	2.39	16.27
46	2004	2.60	16.92
47	2005	3.66	17.58
48	2006	4.72	18.24
49	2007	5.79	18.89

File output_interpolasi.dat

output_interpolasi.dat

```
#Tahun  Pop_Interp  Net_Interp
2005    230.97  3.66
2006    234.02  4.72
2015    261.70  22.21
2016    264.52  27.27

```

Link Github:

https://github.com/BonifasiusRaditya/komnum_pemogramanA/blob/main/cb.ipynb?short_path=a16b9e2