

Predicting Life Expectancy in Diverse Countries Using Multivariate Linear Regression

Dosen Pengampu : Alfi Fadliana S.SI., M.Stat



Disusun Oleh:

Wahyu Ikbil Maulana

(3323600002)

Sains Data Terapan B

POLITEKNIK ELEKTRONIKA NEGERI SURABAYA

DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER

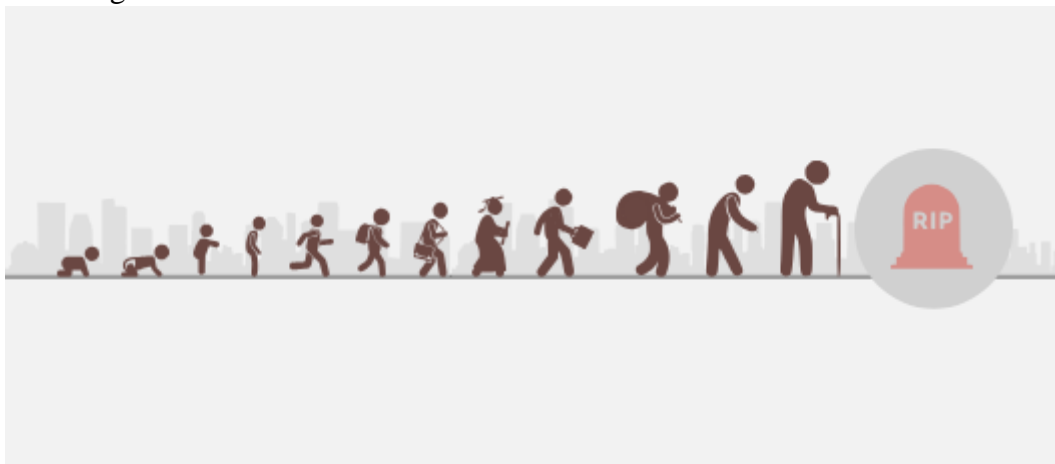
PROGRAM STUDI SAINS DATA TERAPAN

DAFTAR ISI

COVER.....	1
DAFTAR ISI.....	2
Pendahuluan.....	2
Tinjauan Pustaka.....	3
DISKUSI DAN ANALISIS	4
KESIMPULAN.....	8
DAFTAR PUSTAKA	9

Pendahuluan

Latar Belakang



Life expectancy atau harapan hidup adalah suatu penilaian dasar dari kesehatan dan kesejahteraan populasi dari suatu negara, yang berfungsi sebagai indikator penting untuk menilai efektivitas sistem perawatan kesehatan dan pembangunan ekonomi.

Penting untuk mengidentifikasi variabel-variabel yang mempengaruhi dan memprediksi angka harapan hidup penduduk suatu negara agar dapat mengembangkan strategi yang tepat untuk meningkatkan kualitas dan kinerja sistem perawatan kesehatan dan dengan demikian meningkatkan angka harapan hidup.

Penelitian dilakukan untuk memperkirakan angka harapan hidup berfokus pada pemodelan angka harapan hidup dengan menggunakan multivariate linear regression atau multiple linear regression berdasarkan tren kesehatan dan kematian yang diamati pada populasi, serta faktor sosial, ekonomi, dan lingkungan di suatu negara.

Dalam penelitian ini, kami bertujuan untuk membangun linear regresi berganda dengan menggunakan teknik sederhana untuk mengidentifikasi variabel-variabel yang berpengaruh terhadap angka harapan hidup saat lahir.

Tujuan Penelitian

- Mengetahui faktor faktor yang mempengaruhi harapan hidup suatu negara
- Mengetahui solusi dan rekomendasi bagaimana meningkatkan harapan hidup di negara

- Mengetahui model regresi terbaik dari analisis pengaruh variabel-variabel yang diduga terhadap harapan hidup di berbagai negara tahun 2019.

Batasan Masalah

Batasan masalah yang digunakan oleh peneliti adalah harapan hidup dari 168 Negara pada tahun 2019 yang diperoleh dari WHO.

Tinjauan Pustaka

Sustainable Development Goals (SDGs) atau Tujuan Pembangunan Berkelanjutan adalah serangkaian tujuan yang dibuat oleh Perserikatan Bangsa-Bangsa (PBB) sebagai panduan bagi seluruh negara anggota untuk mencapai pembangunan berkelanjutan. SDGs disepakati oleh 190 negara dan disahkan melalui sidang umum PBB pada 25 September 2015 di New York, Amerika Serikat. Agenda pembangunan global ini berlaku mulai dari tahun 2015 hingga 2030. SDGs terdiri dari 17 tujuan yang saling terkait dan saling mempengaruhi satu sama lain.



Tujuan SDGs adalah menjaga peningkatan kesejahteraan ekonomi masyarakat secara berkesinambungan, menjaga keberlanjutan kehidupan sosial masyarakat, menjaga kualitas lingkungan hidup serta pembangunan yang inklusif dan terlaksananya tata kelola yang mampu menjaga peningkatan kualitas kehidupan dari satu generasi ke generasi berikutnya. Beberapa agenda MDGs yang belum tercapai akan dilanjutkan dalam pelaksanaan pencapaian SDGs hingga tahun 2030. Target yang saya spesifik ingin saya capai adalah target 3.4 dengan salah satu tujuan sub indikatornya adalah meningkatkan kesehatan mental dan kesejahteraan dengan meningkatkan angka harapan hidup.

Dalam penelitian ini, kami menyajikan pendekatan inovatif untuk meramalkan harapan hidup di berbagai negara. Dengan menggunakan analisis regresi berganda saya menemukan estimasi nilai koefisien regresi untuk setiap satuan variabel independen, dengan asumsi variabel lainnya tetap konstan. Selain itu, saya juga menguji signifikansi statistik dari koefisien regresi, yang menunjukkan apakah ada hubungan yang signifikan antara variabel independen dan variabel dependen atau tidak.

Regresi Linier Berganda

Analisis regresi berganda juga dapat digunakan untuk memprediksi nilai variabel harapan hidup berdasarkan nilai-nilai variabel independen yang diketahui, dengan menggunakan persamaan regresi yang telah dibentuk.

Analisis linear berganda adalah sebuah metode statistik yang digunakan untuk menguji hubungan antara satu variabel dependen dengan dua atau lebih variabel independen. Tujuan dari analisis ini adalah untuk memprediksi nilai variabel dependen berdasarkan nilai-nilai variabel independen, serta untuk mengetahui seberapa besar pengaruh variabel independen terhadap variabel dependen. Analisis linear berganda juga dapat digunakan untuk menguji hipotesis tentang koefisien regresi, yang menunjukkan kemiringan garis regresi untuk setiap variabel independen. Rumus umum dari analisis linear berganda adalah:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n + e$$

Dimana:

Y adalah variabel dependen

a adalah konstanta atau titik potong garis regresi

b₁, b₂, ..., b_n adalah koefisien regresi untuk variabel independen X₁, X₂, ..., X_n

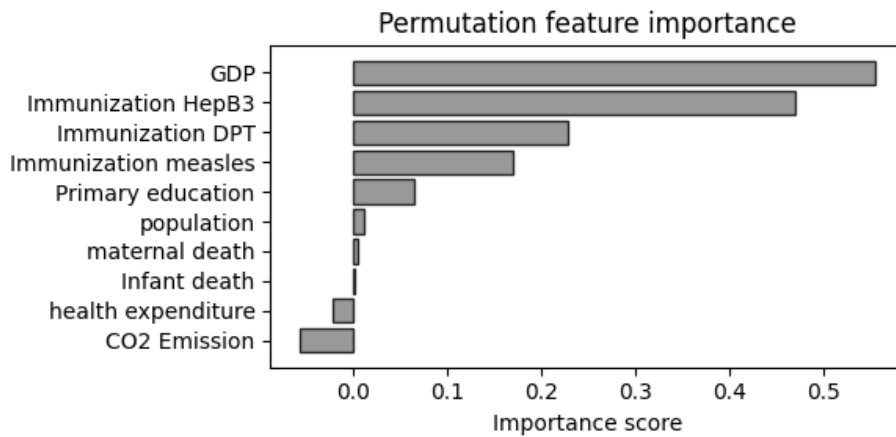
e adalah galat atau residu

CODE DAN ANALISIS

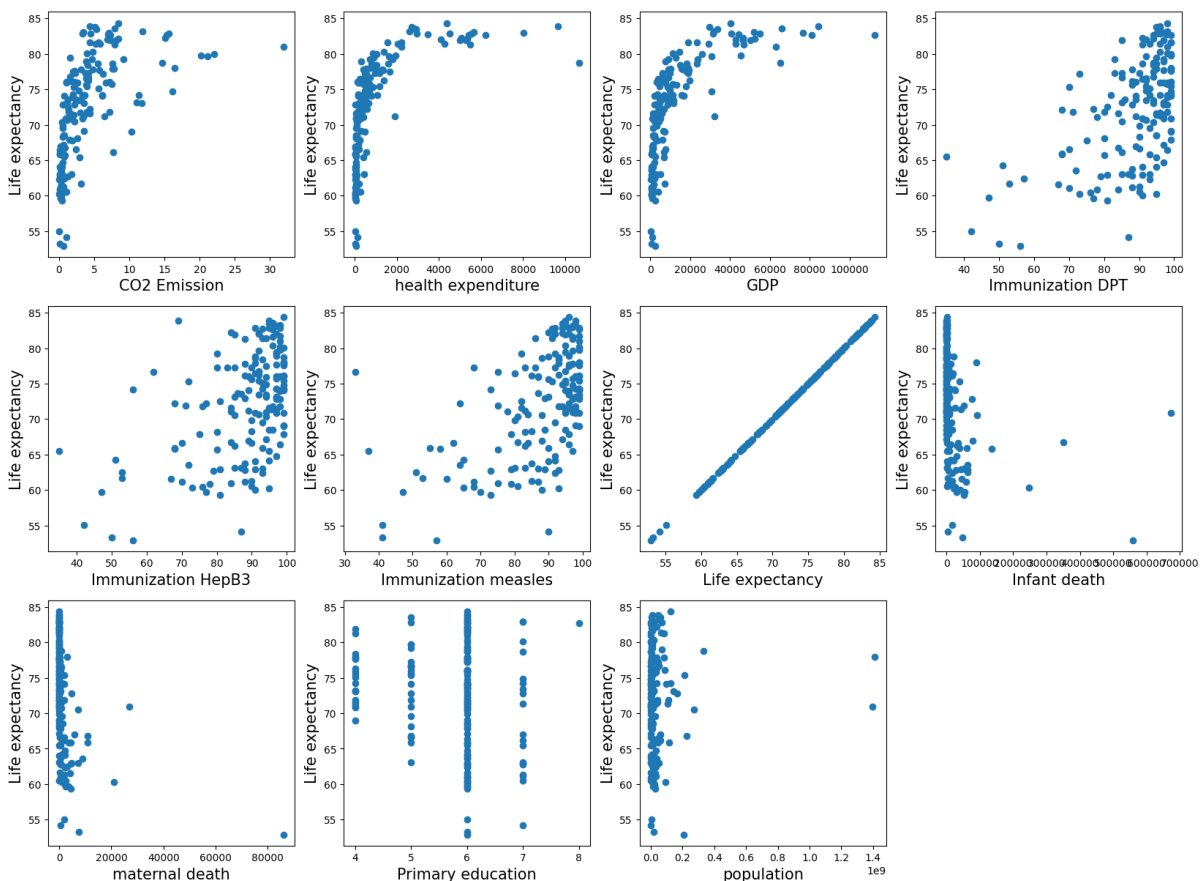
```
#import library

import pandas as pd
import statsmodels.api as sm
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
```

	country	CO2 Emission	health expenditure	GDP	Immunization DPT	Immunization HepB3	Immunization measles	Life expectancy	Infant death	maternal death	Primary education	population
0	Afghanistan	0.297564	74.234104	500.522981	72.0	72.0	64.0	63.565	62760.0	8900.0	6.0	38972230
1	Algeria	3.994402	250.560776	4021.983608	91.0	91.0	80.0	76.474	20101.0	760.0	5.0	43451666
2	Angola	0.753638	69.379341	2142.238757	57.0	53.0	51.0	62.448	63492.0	2900.0	6.0	33428486
3	Antigua and Barbuda	5.525582	801.543823	18187.779712	95.0	99.0	93.0	78.691	6.0	0.0	7.0	92664
4	Argentina	3.742030	958.523499	9963.674231	83.0	83.0	86.0	77.284	5086.0	220.0	6.0	45376763
...
163	Uzbekistan	3.504663	98.587418	1795.201768	96.0	96.0	98.0	71.344	10886.0	240.0	4.0	34232050
164	Vanuatu	0.548285	101.974411	3076.589886	90.0	90.0	80.0	69.877	192.0	8.0	6.0	311685
165	Viet Nam	3.567848	173.650452	3491.091279	89.0	89.0	95.0	74.093	25197.0	2000.0	5.0	96648685
166	Zambia	0.414336	67.457771	1268.120941	88.0	88.0	93.0	62.793	27246.0	840.0	7.0	18927715
167	Zimbabwe	0.663338	54.808506	1421.868596	90.0	90.0	85.0	61.292	17852.0	1900.0	7.0	15669666



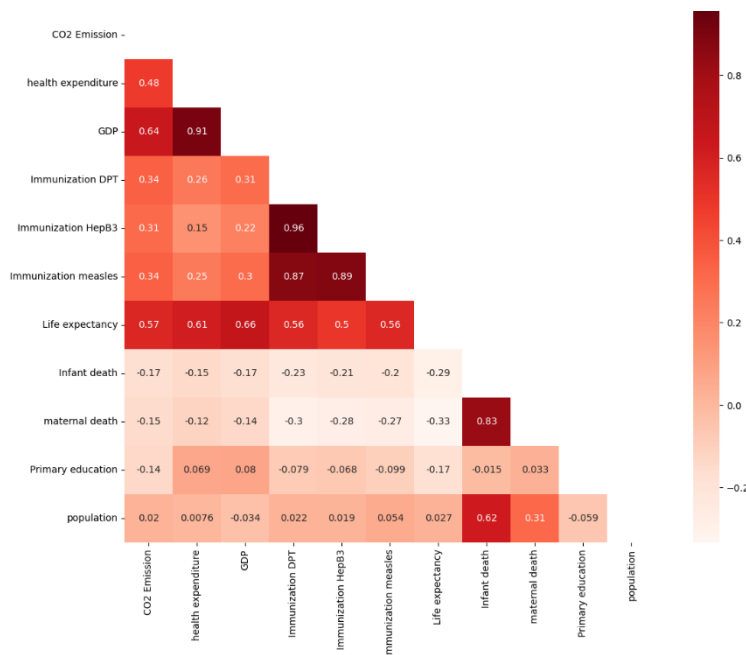
```
plt.figure(figsize=(20,20),dpi=100)
for index, feature_name in enumerate(cols):
    plt.subplot(4,4,index+1)
    plt.scatter(cols[feature_name],data['Life expectancy'])
    plt.ylabel('Life expectancy',size=15)
    plt.xlabel(feature_name,size=15)
```



```

cols = data.select_dtypes(exclude=['object'])
def correlation_matrix(dataframe, x, y):
    corr = dataframe.corr()
    f, ax = plt.subplots(figsize=(x, y))
    mask = np.triu(np.ones_like(corr))
    sns.heatmap(corr, annot=True, mask = mask, cmap="Reds", square=True)
    return ax
correlation_matrix(cols, 14, 10)

```



```

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3,
random_state= 24)
lr = LinearRegression()
lr.fit(x_train, y_train)

print('The r2 score : ', lr.score(x_train, y_train))

```

The r2 score : 0.7104321432870406

```
# Menampilkan persamaan
equation = "y = {:.4f}".format(intercept)

for i, coef in enumerate(coefficients):
    equation += " + {:.8f} * X{}".format(coef, i+1)

print("Persamaan Multivariate Linear Regression:")
print(equation)
```

Persamaan Multivariate Linear Regression:

$$y = 62.6074 + 0.19570180 * X1 + 0.00013036 * X2 + 0.00024693 * X3 + -0.21685745 * X4 + 0.23862620 * X5 + 0.12555256 * X6 + -0.00001059 * X7 + -0.00009389 * X8 + -1.22036529 * X9 + 0.00000001 * X10$$

OLS Regression Results							
Dep. Variable:	y		R-squared:	0.710			
Model:	OLS		Adj. R-squared:	0.683			
Method:	Least Squares		F-statistic:	26.01			
Date:	Thu, 30 Nov 2023		Prob (F-statistic):	3.08e-24			
Time:	03:12:40		Log-Likelihood:	-328.40			
No. Observations:	117		AIC:	678.8			
Df Residuals:	106		BIC:	709.2			
Df Model:	10						
Covariance Type:	nonrobust						
	coef	std err	t	P> t	[0.025	0.975]	
const	62.6074	4.672	13.401	0.000	53.345	71.870	
CO2 Emission	0.1957	0.135	1.446	0.151	-0.073	0.464	
health expenditure	0.0001	0.001	0.144	0.886	-0.002	0.002	
GDP	0.0002	9.75e-05	2.532	0.013	5.36e-05	0.000	
Immunization DPT	-0.2169	0.177	-1.223	0.224	-0.568	0.135	
Immunization HepB3	0.2386	0.168	1.417	0.159	-0.095	0.573	
Immunization measles	0.1256	0.071	1.776	0.079	-0.015	0.266	
Infant death	-1.059e-05	1.07e-05	-0.986	0.326	-3.19e-05	1.07e-05	
maternal death	-9.389e-05	9.14e-05	-1.028	0.306	-0.000	8.72e-05	
Primary education	-1.2204	0.483	-2.528	0.013	-2.178	-0.263	
population	5.612e-09	3.14e-09	1.786	0.077	-6.19e-10	1.18e-08	
Omnibus:	1.488	Durbin-Watson:	2.224				
Prob(Omnibus):	0.475	Jarque-Bera (JB):	1.279				
Skew:	-0.256	Prob(JB):	0.528				
Kurtosis:	3.005	Cond. No.	2.29e+09				

- R-squared: Nilai R-squared sebesar 0.71 menunjukkan bahwa 71% variasi dalam variabel dependen (y) dapat dijelaskan oleh variabel independen dalam model ini.
- Adjusted R-squared: Nilai Adjusted R-squared sebesar 0.683 menunjukkan bahwa setelah penyesuaian untuk jumlah prediktor dalam model, sekitar 68.3% variasi dalam variabel dependen masih dapat dijelaskan.
- F-statistic: Nilai F-statistic sebesar 26.01 dan Prob (F-statistic) sebesar 3.08e-24 menunjukkan bahwa model secara keseluruhan signifikan secara statistik.
- CO2 Emission: Koefisien 0.1957 menunjukkan bahwa untuk setiap peningkatan satu unit dalam emisi CO2, variabel dependen akan meningkat sebesar 0.1957 unit, asumsikan variabel

lain tetap. Namun, nilai p-value 0.151 lebih besar dari 0.05, yang berarti bahwa efek ini mungkin signifikan secara statistik.

- Health Expenditure: Koefisien 0.0001 menunjukkan bahwa untuk setiap peningkatan satu unit dalam pengeluaran kesehatan, variabel dependen akan meningkat sebesar 0.0001 unit, asumsikan variabel lain tetap. Namun, nilai p-value 0.254 lebih besar dari 0.05, yang berarti bahwa efek ini mungkin tidak signifikan secara statistik.
- GDP: Koefisien 0.0001 menunjukkan bahwa untuk setiap peningkatan satu unit dalam GDP, variabel dependen akan meningkat sebesar 0.0001 unit, asumsikan variabel lain tetap. Nilai p-value 0.886 lebih dari 0.05, yang berarti bahwa efek ini tidak signifikan secara statistik.
- Primary Education: Koefisien -1.3736 menunjukkan bahwa untuk setiap peningkatan satu unit dalam pendidikan dasar, variabel dependen akan berkurang sebesar 1.3736 unit, asumsikan variabel lain tetap. Nilai p-value 0.002 kurang dari 0.05, yang berarti bahwa efek ini signifikan secara statistik.
- Variabel lain seperti “Immunization DPT”, “Immunization HepB3”, “Immunization measles”, “Infant death”, dan “maternal death” memiliki nilai p-value lebih besar dari 0.05, yang berarti bahwa efek mereka mungkin tidak signifikan secara statistik.

KESIMPULAN

Salah satu cara untuk menganalisa ekspektasi hidup suatu negara adalah dengan menggunakan faktor-faktor seperti pendapatan per kapita, tingkat pendidikan, akses kesehatan, dan kualitas lingkungan. Berdasarkan hasil analisis yang saya lakukan, didapat menarik beberapa kesimpulan berikut:

- Ekspektasi hidup cenderung lebih tinggi di negara yang tingkat GDP yang tinggi. Indikator ekonomi berpengaruh positif terhadap seberapa lama manusia dapat bertahan hidup.
- CO2 Emission adalah faktor yang berpengaruh negatif terhadap ekspektasi hidup, karena tingkat emisi karbon yang tinggi dapat mengakibatkan banyak sumber penyakit, maka cenderung dapat mengurangi umur hidup manusia. Negara-negara dengan tingkat polusi yang lebih tinggi biasanya memiliki angka harapan hidup yang lebih rendah.
- Imunisasi adalah faktor yang berpengaruh positif terhadap ekspektasi hidup, karena imunisasi sejak kecil dapat meningkatkan imun tubuh hingga usia lanjut, Negara-negara yang warganya rajin melakukan imunisasi memiliki angka harapan hidup yang tinggi.
- Health expenditure/spending adalah faktor yang berpengaruh positif terhadap ekspektasi hidup, karena health spending dapat mencerminkan komitmen dan prioritas negara dalam mengalokasikan sumber daya untuk kesehatan. Negara-negara dengan health spending yang lebih tinggi biasanya memiliki angka harapan hidup yang lebih tinggi.
- Kualitas lingkungan juga berpengaruh terhadap ekspektasi hidup, karena lingkungan yang bersih dan sehat dapat mengurangi paparan terhadap polusi, racun, dan faktor-faktor stres lainnya. Negara-negara dengan kualitas lingkungan yang lebih baik biasanya memiliki angka kematian akibat penyakit pernapasan dan kanker yang lebih rendah, serta angka harapan hidup yang lebih tinggi.

Dengan demikian, Anda dapat menyimpulkan bahwa ekspektasi hidup suatu negara dipengaruhi oleh berbagai faktor yang saling berkaitan, dan untuk meningkatkan ekspektasi

hidup, negara-negara perlu berinvestasi dalam pembangunan ekonomi, sosial, dan lingkungan yang berkelanjutan.

DAFTAR PUSTAKA

https://en.wikipedia.org/wiki/Sustainable_Development_Goals

<https://sdgs.bappenas.go.id/sekilas-sdgs/>

<https://sdgs.bappenas.go.id/tentang/>

<https://instiki.ac.id/2023/05/02/mengenal-sustainable-development-goals-sdgs-atau-tujuan-pembangunan-berkelanjutan/>

<https://bappeda.jogjaprovo.go.id/dataku/sdgs>

<https://bakrie.ac.id/articles/397-sustainable-development-goals-adalah-point-penting-untuk-diketahui-gen-z-simak-selengkapnya.html>

<https://sdgs.bappenas.go.id/tujuan-3/>

<https://ieeexplore.ieee.org/document/9777372>

<https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-8408-3>