

# **Analisis Biaya Kesehatan Terhadap Tingkat Kematian dan Indikasi Terjangkitnya TBC**

**Tugas Kelompok**

**Kelas Penambangan Data Gab 1**

**Taufik Fathurahman [1301160790]**

**Haddad Alwi Yafie [1301162756]**

**Refka Muhammad Furqon [1301162748]**



**Program Studi Sarjana Informatika**

**Fakultas Informatika**

**Universitas Telkom**

**Bandung 2019**

## **A. Latar Belakang Masalah**

Dalam tugas besar ini terdapat dataset tentang kesehatan, yang terdiri dari 12 dataset, yang diantaranya adalah data *current health expenditure*, *health*, *human development index*, *life expectancy at birth*, *life expectancy at birth female*, *life expectancy at birth male*, *life expectancy index*, *malaria incidence*, *mortality rate infant*, *mortality rate male adult*, *mortality rate under-five*, dan *tuberculosis incidence*. Dari data-data tersebut masih belum ada informasi yang dapat diambil, karena masih belum tahu korelasi yang didapat dari data-data tersebut. Kemudian diambil tiga data yang dicoba untuk menemukan korelasi antar data sehingga dapat menghasilkan informasi, tiga data tersebut yaitu *mortality rate male adult*, *mortality rate under-five*, dan *tuberculosis incidence*. Dari ketiga data tersebut ternyata ditemukan korelasinya, dimana ketika data *mortality rate male adult* dan *mortality rate under-five* tinggi/meningkat maka data *tuberculosis incidence* juga meningkat. Sebaliknya ketika data *mortality rate male adult* dan *mortality rate under-five* menurun maka data *tuberculosis incidence* juga menurun. Setelah ditemukan korelasinya maka diputuskan untuk menggali lebih dalam informasi yang akan diambil.

Untuk memperoleh informasi dari dataset *mortality rate male adult*, *mortality rate under-five*, dan *tuberculosis incidence*, akan digunakan metode *Gaussian Mixture Model Clustering*. Proses dari clustering akan menghasilkan dua kelompok negara, yaitu kelompok negara dengan tingkat kematian tinggi dan rendah. Setelah didapat informasi hasil clustering, hasilnya akan dicari relasinya dengan tingkat pengeluaran biaya kesehatan. Dari hasil relasi yang terbentuk, diharapkan mendapat informasi bagaimana biaya kesehatan dapat menekan tingkat kematian yang terjadi.

## **B. Tujuan**

Tujuan dari analisis data kesehatan ini yaitu mendapatkan pola dan solusi dari tingkat kematian usia dibawah lima tahun, tingkat kematian pada pria dewasa, dan indikasi TBC terhadap biaya pengeluaran kesehatan.

C. Deskripsi Data

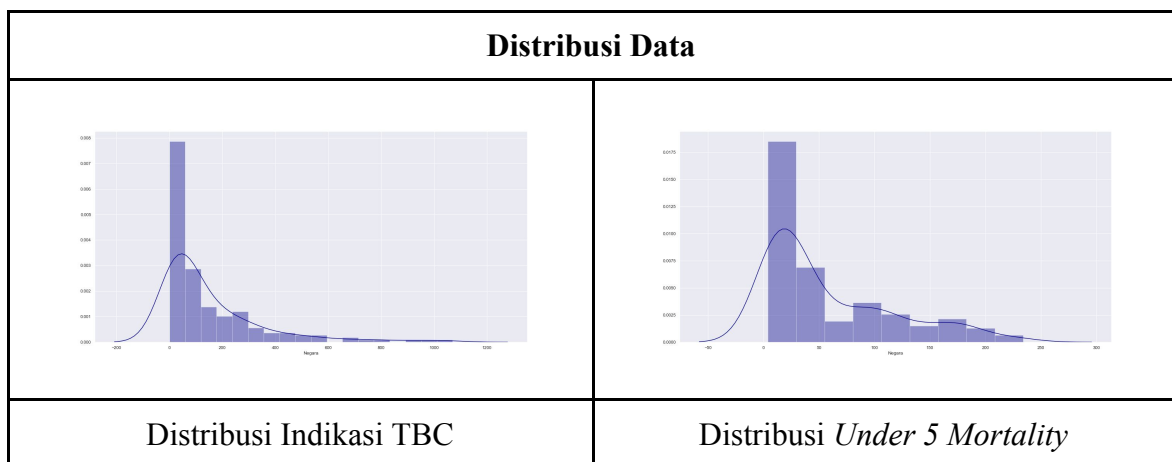
Tabel 1: Karakteristik Data

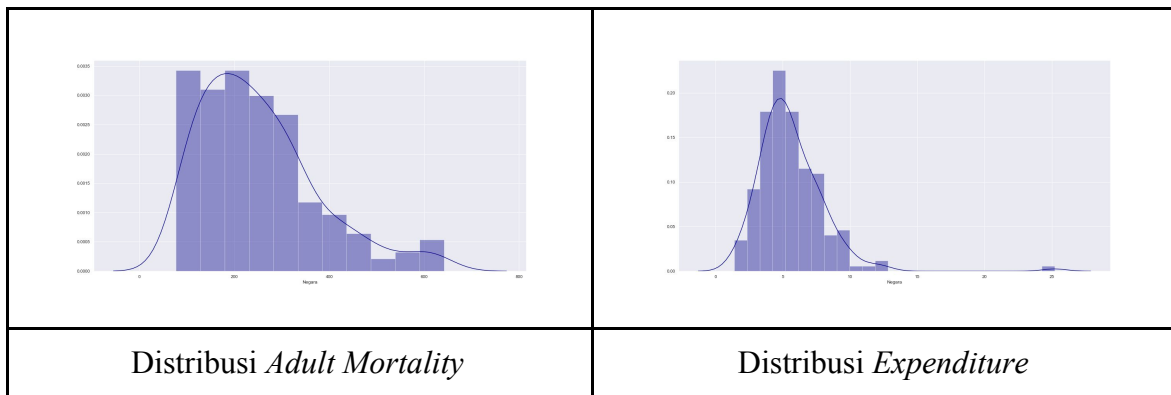
Dataset	Jumlah Baris	Jumlah Kolom	Kolom	Tipe Data Setiap Kolom	Missing value
TBC Idikasi	194	11	HDI Rank (2017)	integer	6
			Country	String	0
			2000	Float	4
			2005	Float	1
			2010	Float	1
			2011	Float	0
			2012	Float	0
			2013	Float	0
			2014	Float	0
			2015	Float	0
			2016	Float	0
Under 5 Mortality	193	13	HDI Rank (2017)	integer	6
			Country	String	0
			1990	Float	0
			1995	Float	0
			2000	Float	0
			2005	Float	0

			2010	Float	0
			2011	Float	0
			2012	Float	0
			2013	Float	0
			2014	Float	0
			2015	Float	0
			2016	Float	0
Adult Mortality	188	13	HDI Rank (2017)	integer	2
			Country	String	0
			1990	Float	1
			1995	Float	3
			2000	Float	0
			2005	Float	3
			2010	Float	3
			2011	Float	4
			2012	Float	5
			2013	Float	5
			2014	Float	8
			2015	Float	31

			2016	Float	37
Expenditure	188	10	HDI Rank (2017)	integer	4
			Country	String	0
			2000	Float	5
			2005	Float	2
			2010	Float	2
			2011	Float	2
			2012	Float	1
			2013	Float	2
			2014	Float	2
			2015	Float	2

Tabel 2: Distribusi Data (x, y: index, negara)





Dari Tabel 1 dan 2, didapatkan hasil bahwa dataset yang akan digunakan belum memiliki kualitas yang baik. Alasan pertama dari kualitas dataset yang belum baik adalah, keempat dataset tidak memiliki jumlah kolom dan baris yang sama, sehingga akan ada kolom atau baris yang tidak memiliki relasi. Alasan kedua adalah masih banyak terdapat *missing value* pada keempat dataset. Dan untuk alasan terakhir adalah distribusi dari *dataset* yang belum normal. Distribusi normal dibutuhkan disini karena algoritma *clustering* yang akan digunakan yaitu *gaussian mixture model* menggunakan *step* ekspektasi dan maksimasi.

#### D. Praproses

Praproses yang diterapkan untuk menangani *dataset* yang digunakan terdiri dari beberapa step berikut ini:

1. Menyeleksi kolom dan baris

Menyeleksi kolom dan baris yang ada pada semua dataset yang digunakan, dengan kata lain melakukan operasi intersection. Ini bertujuan agar tidak ada kolom dan baris yang tidak memiliki relasi.

**Hasil** : jumlah baris dan kolom untuk keempat *dataset* = (182, 9)

2. Menangani *missing value*

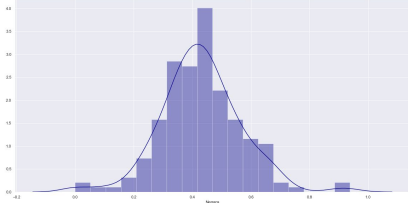
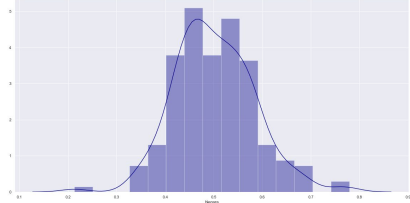
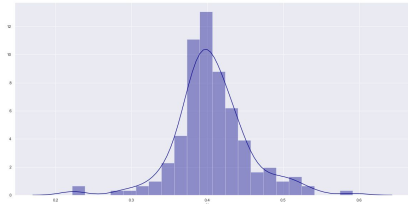
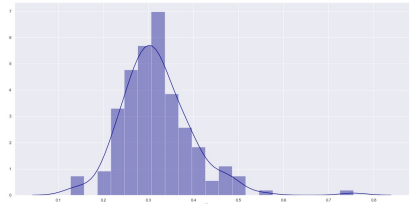
Tahapan menangani *missing value* dilakukan dengan mengganti nilai yang *missing* dengan menggunakan *mean*. *Mean* dipilih karena dataset yang dipilih tidak memiliki banyak *outlier*.

### 3. Normalisasi Data

Normalisasi data dilakukan dengan cara *zero center*, yaitu mengurangi data dengan *mean*, lalu membaginya dengan standar deviasi. Penerapan normalisasi dilakukan karena algoritma *clustering* yang akan digunakan adalah *gaussian mixture model*, dimana algoritma ini menerapkan *step* ekspektasi dan maksimasi.

**Hasil :**

Tabel 3: Hasil normalisasi (x, y: index, negara)

	
Indikasi TBC	<i>Under 5 Mortality</i>
	
<i>Adult Mortality</i>	<i>Expenditure</i>

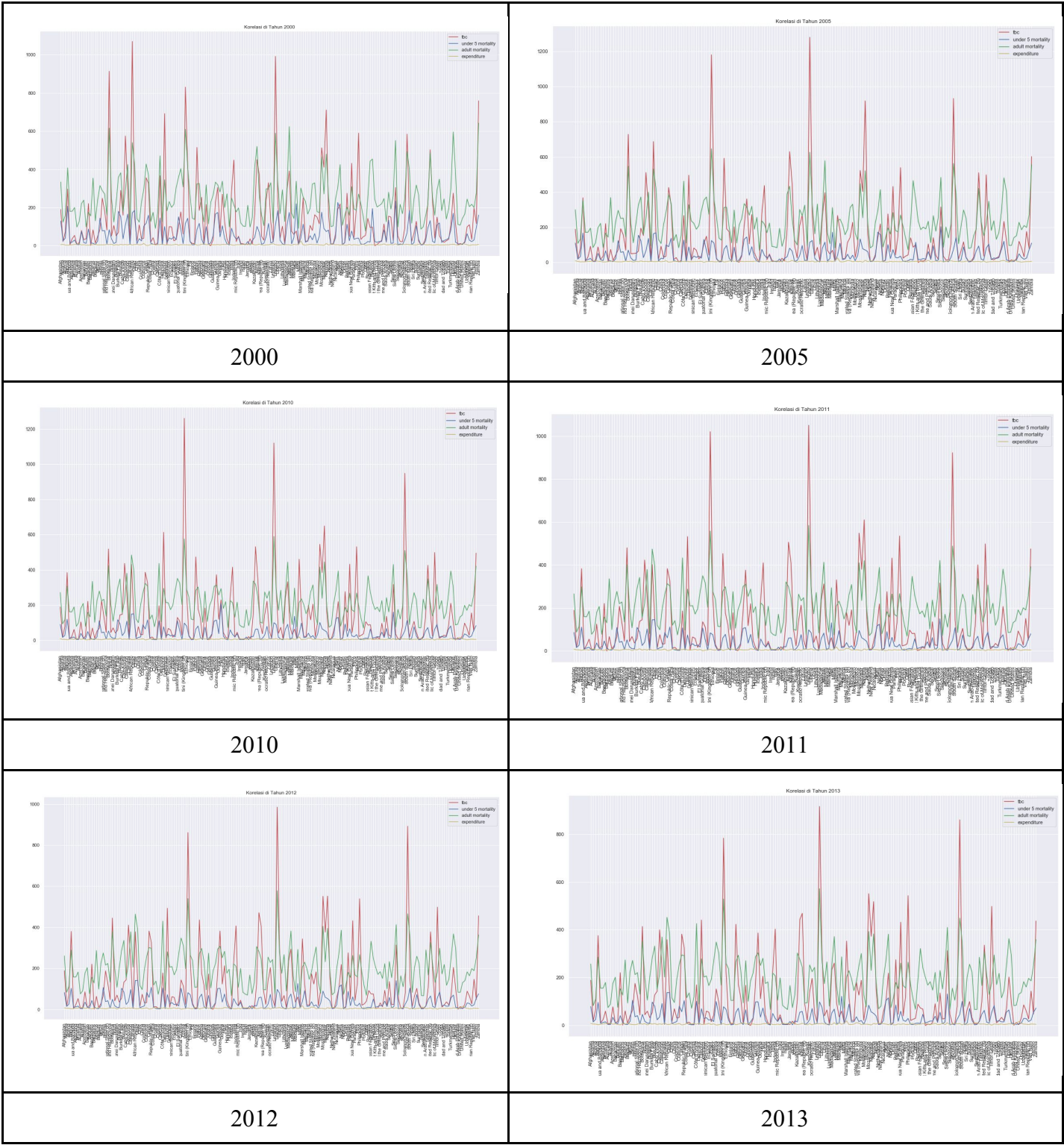
### 4. Mencari relasi antar *dataset*

Pencarian relasi dilakukan untuk menentukan algoritma yang cocok untuk diterapkan pada dataset yang dimiliki. Pencarian relasi diterapkan terhadap 3

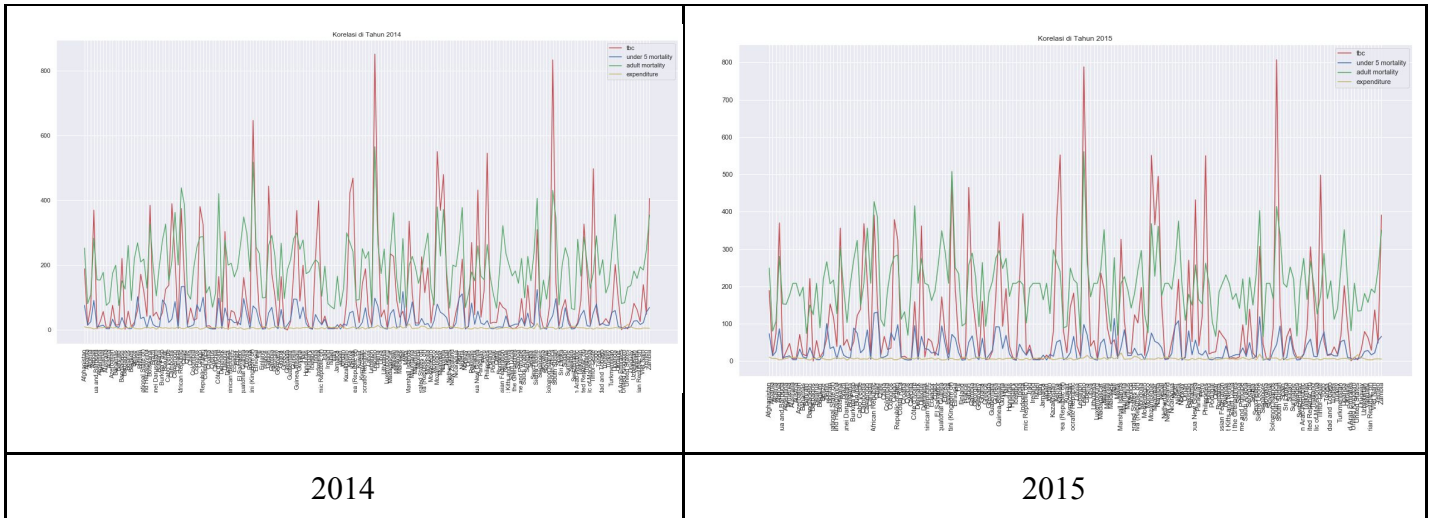
dataset, yaitu *mortality rate male adult*, *mortality rate under-five*, dan *tuberculosis incidence*.

Hasil :

Tabel 4: Tabel relasi pertahun (x, y: index, negara)





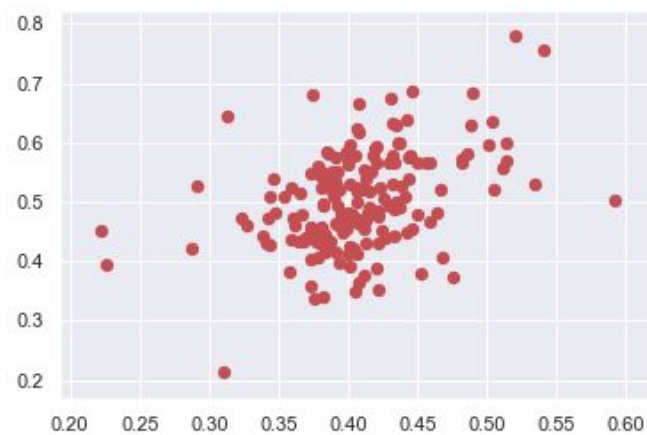


Dari Tabel 4, didapatkan hasil bahwa ketiga dataset memiliki korelasi yang kuat, ditunjukkan dengan bila indikasi TBC (warna merah) tinggi maka tingkat kematian *under 5* dan *adult* (biru, dan hijau) ikut meningkat juga. Relasi ini yang akan menjadi landasan dalam menentukan algoritma pada tahapan selanjutnya.

#### E. Analisis Pemilihan Algoritma

Analisis dilakukan dengan memperhatikan beberapa hal berikut ini:

1. Semua dataset yang dipilih tidak memiliki label, sehingga algoritma yang cocok digunakan merupakan algoritma yang ada pada *unsupervised learning*.
2. Pengelompokan data dibutuhkan untuk mengelompokkan tingkat kematian menjadi dua kelompok, yaitu tinggi dan rendah.
3. Tipe data dari kolom yang akan dijadikan *features* untuk dilakukan pengelompokan merupakan *float*.
4. Sebaran data yang kurang beraturan, seperti pada Gambar 1 dibawah.



Gambar 1. Plot sampel sebaran data salah pada tahun 2000

Sebaran data bersifat acak, sehingga *gaussian mixture model* kami anggap lebih cocok untuk menanganinya.

Dari pertimbangan tiga hal diatas, maka dipilih algoritma *Clustering Gaussian Mixture Model* menjadi algoritma yang dipilih, karena dapat menangani 4 point diatas. Mulai dari dataset yang tidak memiliki label, sampai sebaran data yang acak.

#### F. Analisis Penentuan Parameter

Parameter yang digunakan dalam melakukan *clustering* dengan *gaussian mixture model* adalah sebagai berikut :

1. Jenis kovarian yaitu 'full', ini akan menentukan bentuk dari *cluster* yang dibentuk.
2. *Covariance threshold*, menggunakan nilai  $1e-3$ . Digunakan untuk memberhentikan iterasi dari operasi ekspektasi dan maksimasi ketika dibawah nilai *threshold*.
3. Regularization sebesar  $1e-6$ , yang bertujuan untuk memastikan matriks kovarian selalu bernilai positif.
4. Maksimum iterasi adalah 100. Merupakan berapa banyak iterasi dari proses *ekspektasi maksimasi* akan dilakukan.
5. Banyaknya cluster adalah 2, karena ingin mengelompokkan data kedalam 2 kelompok yaitu *mortality* tinggi dan rendah.

## G. Hasil Percobaan

Proses *clustering* dilakukan pada 183 negara yang ada dari tahun 2000 sampai 2015. Hasil dari proses *clustering* akan ditampilkan pada bentuk tabel, karena untuk memudahkan dalam melakukan proses analisis pertahunnya. Analisis pertahun ini untuk melihat perubahan dari anggota cluster setiap tahunnya.

Tabel 5: Hasil *Clustering*

Country	2000	2005	2010	2011	2012	2013	2014	2015
Afghanistan	1	1	0	1	0	0	0	0
Albania	0	0	0	1	0	1	1	0
Algeria	0	1	0	1	0	0	1	0
Angola	0	1	0	1	0	0	0	0
Antigua and Barbuda	1	1	0	1	1	1	0	0
Argentina	1	1	0	1	0	0	0	0
Armenia	1	1	0	1	0	0	0	0
Australia	0	0	1	0	1	1	1	1
Austria	1	0	1	0	1	1	1	1
Azerbaijan	1	1	0	1	0	0	0	0
Bahamas	1	1	0	1	1	1	0	0
Bahrain	1	1	0	1	0	0	0	0
Bangladesh	1	1	0	1	0	0	0	0
Barbados	0	1	0	1	0	0	0	0
Belarus	1	0	1	0	0	0	1	0
Belgium	1	1	0	1	0	0	1	0
Belize	1	1	0	1	0	0	0	0
Benin	1	1	0	1	0	0	0	0
Bhutan	0	1	0	1	0	0	0	0
Bolivia (Plurinational State of)	1	1	0	1	0	0	0	0
Bosnia and	1	1	0	1	0	0	0	0

<b>Herzegovina</b>							
<b>Botswana</b>	0	0	0	1	1	1	0
<b>Brazil</b>	1	1	0	1	0	0	0
<b>Brunei Darussalam</b>	1	1	0	1	0	0	0
<b>Bulgaria</b>	1	1	0	1	0	0	0
<b>Burkina Faso</b>	1	1	0	1	0	0	0
<b>Burundi</b>	1	1	0	1	0	0	0
<b>Cabo Verde</b>	1	1	0	1	0	0	0
<b>Cambodia</b>	0	1	0	0	0	0	0
<b>Cameroon</b>	1	1	0	1	0	0	0
<b>Canada</b>	0	0	1	0	1	1	1
<b>Central African Republic</b>	1	1	0	1	1	0	0
<b>Chad</b>	1	1	0	1	0	0	0
<b>Chile</b>	1	1	0	1	0	0	0
<b>China</b>	0	1	0	0	0	0	0
<b>Colombia</b>	1	1	0	1	0	0	0
<b>Comoros</b>	1	1	0	1	0	0	0
<b>Congo</b>	0	0	0	1	0	0	1
<b>Congo (Democratic Republic of the)</b>	1	1	0	1	0	0	0
<b>Costa Rica</b>	1	1	0	1	0	0	0
<b>Croatia</b>	1	1	0	1	1	1	0
<b>Cyprus</b>	1	1	0	1	0	0	0
<b>Czechia</b>	1	1	1	0	1	1	1
<b>Côte d'Ivoire</b>	1	1	0	1	0	0	0
<b>Denmark</b>	1	0	1	0	1	1	1
<b>Djibouti</b>	1	1	0	1	0	0	0
<b>Dominica</b>	0	1	1	1	0	1	0
<b>Dominican Republic</b>	1	1	0	1	0	0	0
<b>Ecuador</b>	1	1	0	1	0	0	0

<b>Egypt</b>	1	1	0	1	0	0	0	0
<b>El Salvador</b>	1	1	0	1	0	0	0	0
<b>Equatorial Guinea</b>	1	1	0	1	0	0	0	0
<b>Eritrea</b>	1	0	0	1	0	0	0	0
<b>Estonia</b>	0	0	0	0	0	1	1	0
<b>Eswatini (Kingdom of)</b>	1	0	0	1	0	0	0	0
<b>Ethiopia</b>	0	0	0	0	0	1	1	0
<b>Fiji</b>	0	1	0	1	0	0	0	0
<b>Finland</b>	1	0	0	1	0	0	1	0
<b>France</b>	1	1	0	1	0	0	1	0
<b>Gabon</b>	1	1	0	1	0	0	0	0
<b>Gambia</b>	1	1	0	1	0	0	0	0
<b>Georgia</b>	0	0	0	0	1	1	1	0
<b>Germany</b>	1	1	0	1	0	0	1	0
<b>Ghana</b>	1	1	0	1	0	0	0	0
<b>Greece</b>	1	1	0	1	0	0	0	0
<b>Grenada</b>	0	1	1	1	1	1	0	0
<b>Guatemala</b>	1	1	0	1	0	0	0	0
<b>Guinea</b>	1	1	0	1	0	0	0	0
<b>Guinea-Bissau</b>	1	1	0	1	0	0	0	0
<b>Guyana</b>	1	1	0	1	0	0	0	0
<b>Haiti</b>	1	1	0	0	0	0	0	0
<b>Honduras</b>	1	1	0	1	0	0	0	0
<b>Hungary</b>	1	0	0	1	0	1	1	0
<b>Iceland</b>	0	0	1	0	1	1	1	1
<b>India</b>	1	0	0	1	0	0	0	0
<b>Indonesia</b>	1	1	0	1	0	0	0	0
<b>Iran (Islamic Republic of)</b>	0	0	0	0	1	1	1	0
<b>Iraq</b>	0	1	0	1	0	0	0	0

<b>Ireland</b>	1	0	1	0	1	1	1	1
<b>Israel</b>	0	0	1	0	1	1	1	1
<b>Italy</b>	0	0	1	0	1	1	1	1
<b>Jamaica</b>	1	1	0	1	0	0	0	0
<b>Japan</b>	0	0	1	0	1	1	1	1
<b>Jordan</b>	1	1	0	1	0	0	0	0
<b>Kazakhstan</b>	1	0	0	1	0	0	0	0
<b>Kenya</b>	0	0	1	0	1	1	1	0
<b>Kiribati</b>	1	1	0	1	0	0	0	0
<b>Korea (Republic of)</b>	0	0	0	0	1	1	1	0
<b>Kuwait</b>	1	1	0	1	0	0	0	0
<b>Kyrgyzstan</b>	1	0	0	1	0	0	0	0
<b>Lao People's Democratic Republic</b>	1	1	0	1	0	0	0	0
<b>Latvia</b>	1	0	0	0	0	0	0	0
<b>Lebanon</b>	0	1	0	0	0	1	1	0
<b>Lesotho</b>	1	1	0	1	0	0	0	0
<b>Liberia</b>	1	1	0	1	0	0	0	0
<b>Libya</b>	0	0	0	1	0	1	1	0
<b>Lithuania</b>	1	0	0	1	0	0	0	0
<b>Luxembourg</b>	1	0	1	0	1	1	1	1
<b>Madagascar</b>	1	1	0	1	0	0	0	0
<b>Malawi</b>	0	0	0	0	1	1	1	0
<b>Malaysia</b>	1	1	0	1	0	0	0	0
<b>Maldives</b>	0	1	0	0	0	1	1	0
<b>Mali</b>	1	1	0	1	0	0	0	0
<b>Malta</b>	0	1	0	1	0	1	0	0
<b>Marshall Islands</b>	0	1	0	0	0	1	1	0
<b>Mauritania</b>	1	1	0	1	0	0	0	0
<b>Mauritius</b>	1	1	0	1	0	0	0	0

<b>Mexico</b>	1	1	0	1	0	0	0	0
<b>Micronesia (Federated States of)</b>	1	1	0	1	0	0	0	0
<b>Moldova (Republic of)</b>	1	1	0	1	0	0	0	0
<b>Mongolia</b>	1	1	0	0	0	0	1	0
<b>Montenegro</b>	1	1	1	1	1	1	0	0
<b>Morocco</b>	0	0	1	0	1	1	1	0
<b>Mozambique</b>	1	1	0	1	0	0	0	0
<b>Myanmar</b>	1	1	0	1	0	0	0	0
<b>Namibia</b>	1	0	0	1	0	0	1	0
<b>Nepal</b>	0	1	0	1	0	0	0	0
<b>Netherlands</b>	0	0	1	0	1	1	1	1
<b>New Zealand</b>	0	0	1	0	1	1	1	1
<b>Nicaragua</b>	1	1	0	1	0	0	0	0
<b>Niger</b>	1	1	0	1	0	0	0	0
<b>Nigeria</b>	1	1	0	1	0	0	0	0
<b>Norway</b>	1	0	1	0	1	1	1	1
<b>Oman</b>	1	1	0	1	0	0	1	0
<b>Pakistan</b>	1	1	0	1	0	0	0	0
<b>Panama</b>	1	1	0	1	0	0	0	0
<b>Papua New Guinea</b>	1	1	0	1	0	0	0	0
<b>Paraguay</b>	1	1	0	1	0	0	0	0
<b>Peru</b>	1	1	0	1	0	0	0	0
<b>Philippines</b>	1	1	0	1	0	0	0	0
<b>Poland</b>	1	1	0	1	0	0	1	0
<b>Portugal</b>	1	1	0	1	0	0	1	0
<b>Qatar</b>	1	1	0	1	0	0	0	0
<b>Romania</b>	1	1	0	1	0	0	0	0
<b>Russian Federation</b>	1	0	0	1	0	0	0	0
<b>Rwanda</b>	0	0	0	0	0	1	1	0

<b>Saint Kitts and Nevis</b>	1	1	0	1	0	1	0	0
<b>Saint Lucia</b>	1	1	0	1	0	0	0	0
<b>Saint Vincent and the Grenadines</b>	1	1	0	1	1	1	0	0
<b>Samoa</b>	1	1	0	1	1	1	1	0
<b>Sao Tome and Principe</b>	1	0	0	1	0	0	0	0
<b>Saudi Arabia</b>	1	1	0	1	0	0	0	0
<b>Senegal</b>	1	1	0	1	0	0	0	0
<b>Serbia</b>	1	1	1	1	1	1	0	0
<b>Seychelles</b>	1	1	0	1	0	1	0	0
<b>Sierra Leone</b>	0	1	0	1	0	0	0	0
<b>Singapore</b>	0	1	0	1	0	1	1	0
<b>Slovakia</b>	1	1	0	1	1	1	1	1
<b>Slovenia</b>	1	0	1	0	1	1	1	1
<b>Solomon Islands</b>	0	1	0	1	0	0	1	0
<b>South Africa</b>	1	0	0	1	0	0	0	0
<b>South Sudan</b>	1	1	0	1	0	0	0	0
<b>Spain</b>	1	0	1	0	1	1	1	1
<b>Sri Lanka</b>	1	0	0	1	0	0	0	0
<b>Sudan</b>	1	1	0	1	0	0	0	0
<b>Suriname</b>	1	1	0	1	0	0	0	0
<b>Sweden</b>	0	1	0	1	0	0	0	0
<b>Switzerland</b>	0	0	1	0	1	1	1	1
<b>Syrian Arab Republic</b>	0	0	0	0	1	1	1	0
<b>Tajikistan</b>	1	1	0	1	0	0	0	0
<b>Tanzania (United Republic of)</b>	0	0	0	1	0	1	1	0
<b>Thailand</b>	1	1	0	1	0	0	0	0
<b>The former Yugoslav Republic of Macedonia</b>	1	1	0	1	0	0	0	0



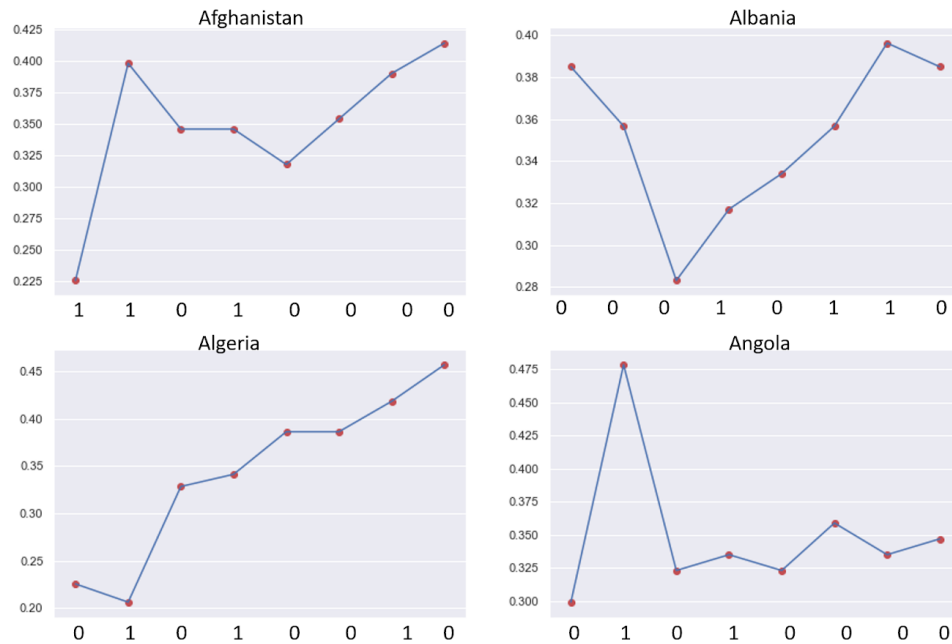
<b>Timor-Leste</b>	0	1	1	0	0	1	1	0
<b>Togo</b>	1	1	0	1	0	0	0	0
<b>Tonga</b>	1	1	0	1	0	1	0	0
<b>Trinidad and Tobago</b>	1	1	0	1	0	0	0	0
<b>Tunisia</b>	1	1	0	1	0	0	0	0
<b>Turkey</b>	1	1	0	1	0	0	0	0
<b>Turkmenistan</b>	1	1	0	1	0	0	0	0
<b>Uganda</b>	0	0	0	0	0	1	1	0
<b>Ukraine</b>	1	0	0	1	0	0	1	0
<b>United Arab Emirates</b>	1	1	0	1	1	1	0	0
<b>United Kingdom</b>	0	0	1	0	1	1	1	1
<b>United States</b>	1	1	0	1	0	0	0	0
<b>Uruguay</b>	1	1	0	1	0	0	0	0
<b>Uzbekistan</b>	1	0	0	1	0	0	0	0
<b>Vanuatu</b>	1	1	0	1	0	0	0	0
<b>Venezuela (Bolivarian Republic of)</b>	1	1	0	1	0	0	0	0
<b>Viet Nam</b>	1	1	0	1	0	0	0	0
<b>Yemen</b>	1	1	0	1	0	0	0	0
<b>Zambia</b>	0	0	0	0	1	1	1	0

Pada Tabel 5 merupakan hasil dari pengelompokkan data menjadi 2 kelompok yang ditandai dengan label 0 dan 1. Tanda 0 berarti tingkat kematian tinggi atau kesehatan buruk, dan 1 merupakan label untuk tingkat kematian rendah atau kesehatan baik. Dari hasil *clustering* ini akan ditampilkan bagaimana tingkat kematian dan kualitas kesehatan sangat bergantung terhadap biaya kesehatan, seperti yang akan disajikan pada Gambar 2.

## H. Interpretasi Model

Model *clustering* yang telah dibuat pada tahapan G akan digunakan dalam menggali informasi bagaimana biaya kesehatan sangat penting untuk mengurangi kematian

dan meningkatkan kesehatan dari warga negaranya. Korelasi ini akan disajikan pada Gambar 2.



Gambar 2 Korelasi antara tingkat kematian dan kesehatan (X axis) dengan biaya kesehatan (Y axis)

Kami mengambil 4 negara teratas untuk dijadikan sampel dalam menganalisis bagaimana biaya kesehatan dapat mengurangi angka kematian dan meningkatkan kualitas kesehatan. Bila kita melihat ketika biaya kesehatan yang dikeluarkan itu tinggi, maka angka kematian cenderung menurun, atau dapat dikatakan kualitas kesehatan menjadi meningkat.

Dari informasi yang didapat, dapat disimpulkan bahwa tujuan pada poin B telah dipenuhi. Dari hasil percobaan ini didapat bahwa peningkatan biaya kesehatan dapat menjadi salah satu solusi yang dapat dipilih untuk menurunkan tingkat kematian dan meningkatkan kesehatan warga. Selain itu juga dapat disimpulkan bahwa indikasi dari terjangkitnya penyakit TBC juga ikut diturunkan.