

# Penerapan Data *Mining* dan Algoritma Pohon Keputusan pada Analisis Faktor Penyebaran HIV/AIDS di Indonesia

...

Muhammad Iqbal Rahmadhan Putra  
Dianisti Saraswati  
Institut Teknologi Bandung

FSM UNDIP, 9 September 2018



# Out line



## Tinjauan Pustaka

- a. Faktor-faktor HIV /AIDS
- b. Data Mining
- c. Algoritma Pohon Keputusan

## Hasil dan Analisis

## Metodologi Penelitian

## Kesimpulan dan Saran

## Pendahuluan

- a. Latar Belakang
- b. Tujuan

# Pendahuluan

# Latar

# Belakang

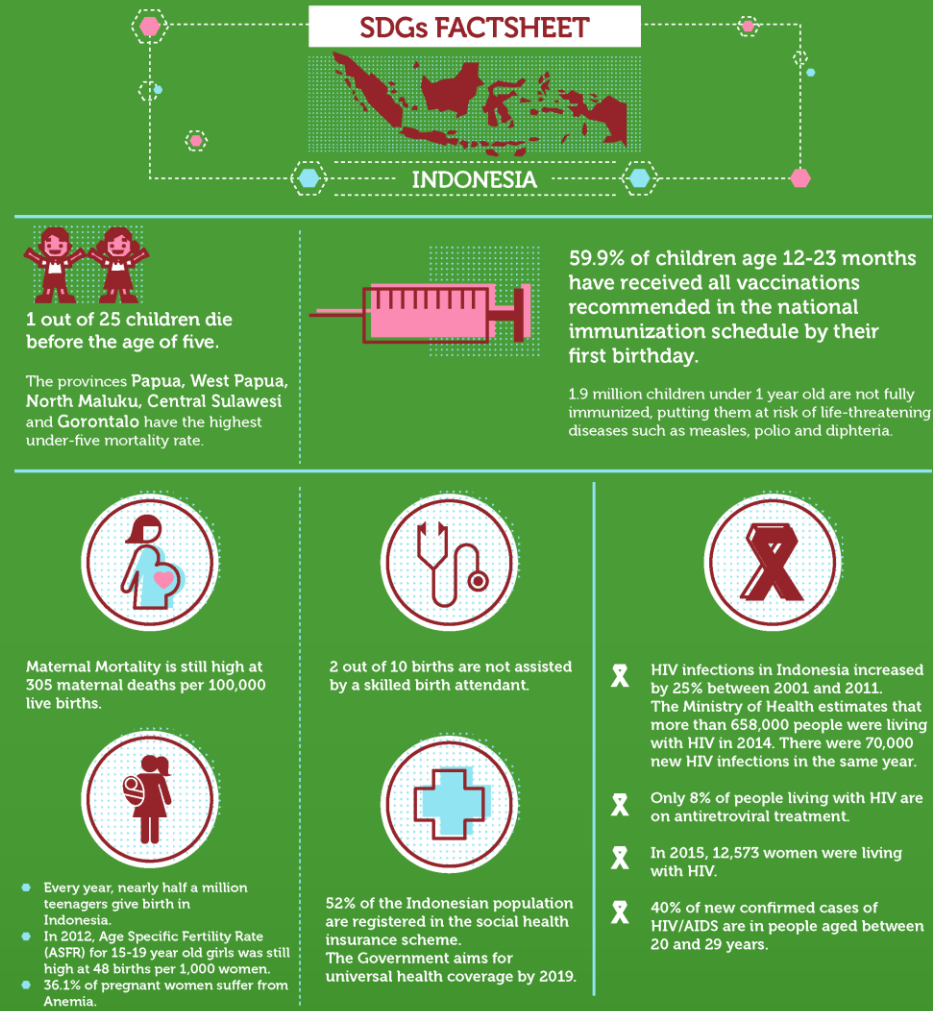
- ✓ Mengakhiri epidemi HIV/AIDS sebagai *Sustainable Development Goal's* yang ketiga
- ✓ Grafik jumlah kasus HIV dan AIDS dari tahun ke tahun
- ✓ Kebutuhan terhadap analisis faktor-faktor penyebaran untuk kebijakan yang efektif

Source : [un.or.id](http://un.or.id)



## GOAL 3 GOOD HEALTH AND WELL-BEING

Ensure healthy lives and promote well-being for all at all ages.



# Pendahuluan

## Tujuan

1. Menentukan implementasi algoritma pohon keputusan pada analisis faktor penyebaran HIV/AIDS di Indonesia
2. Menentukan faktor yang paling berpengaruh terhadap penyebaran HIV/AIDS di Indonesia
3. Menentukan analisis dari aturan yang dibentuk oleh algoritma pohon keputusan

Source : [un.or.id](http://un.or.id)



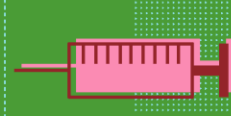
### GOAL 3 GOOD HEALTH AND WELL-BEING

Ensure healthy lives and  
promote well-being for all  
at all ages.



1 out of 25 children die  
before the age of five.

The provinces Papua, West Papua,  
North Maluku, Central Sulawesi  
and Gorontalo have the highest  
under-five mortality rate.



59.9% of children age 12-23 months  
have received all vaccinations  
recommended in the national  
immunization schedule by their  
first birthday.

1.9 million children under 1 year old are not fully  
immunized, putting them at risk of life-threatening  
diseases such as measles, polio and diphtheria.



Maternal Mortality is still high at  
305 maternal deaths per 100,000  
live births.



2 out of 10 births are not assisted  
by a skilled birth attendant.



 HIV infections in Indonesia increased  
by 25% between 2001 and 2011.  
The Ministry of Health estimates that  
more than 658,000 people were living  
with HIV in 2014. There were 70,000  
new HIV infections in the same year.

 Only 8% of people living with HIV are  
on antiretroviral treatment.

 In 2015, 12,573 women were living  
with HIV.

 40% of new confirmed cases of  
HIV/AIDS are in people aged between  
20 and 29 years.

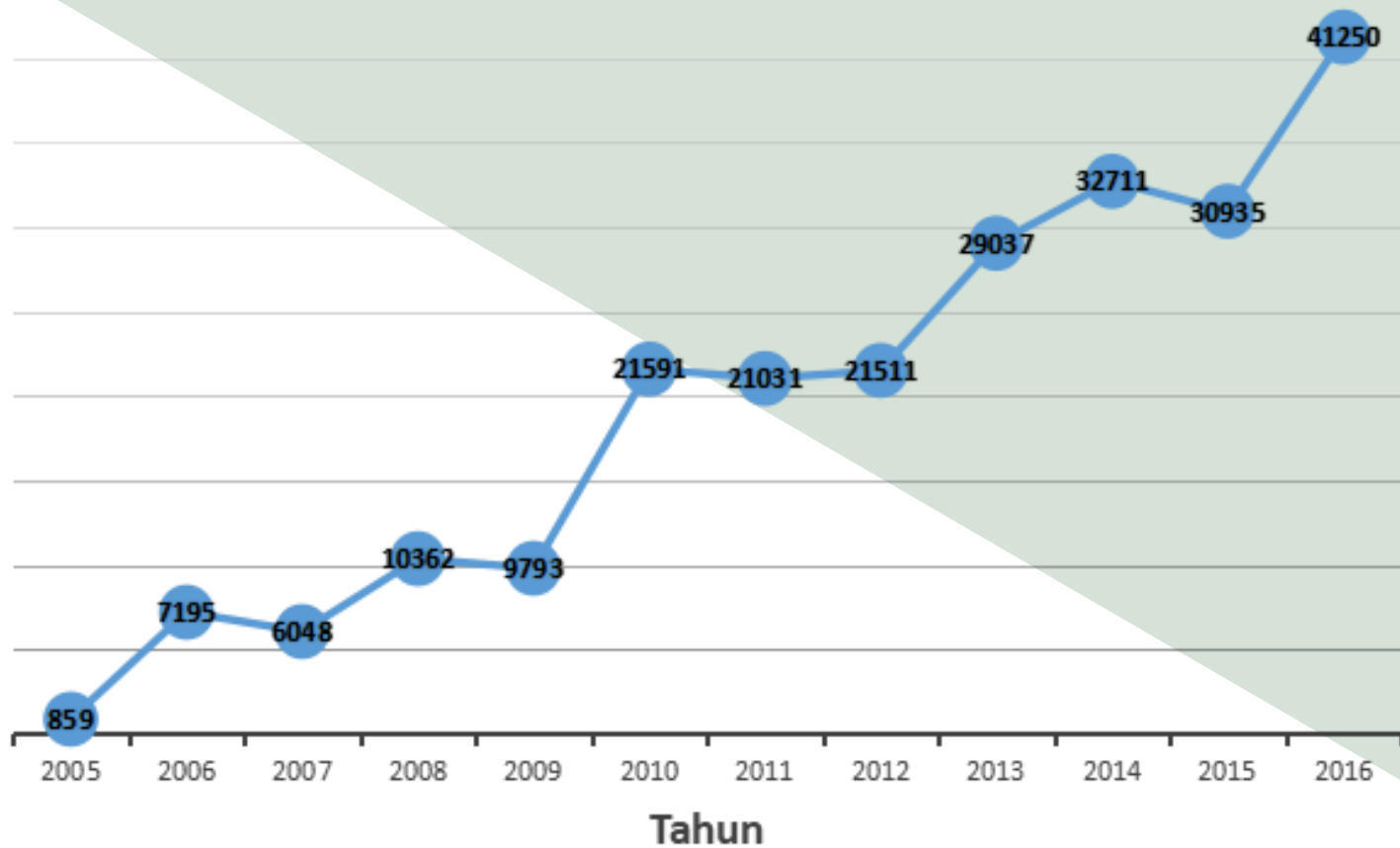


- Every year, nearly half a million  
teenagers give birth in  
Indonesia.
- In 2012, Age Specific Fertility Rate  
(ASFR) for 15-19 year old girls was still  
high at 48 births per 1,000 women.
- 36.1% of pregnant women suffer from  
Anemia.



52% of the Indonesian population  
are registered in the social health  
insurance scheme.  
The Government aims for  
universal health coverage by 2019.

## Jumlah Laporan Kasus HIV



Source:  
Laporan  
Perkembangan  
HIV/AIDS dan  
PIMS di  
Indonesia  
Triwulan  
IV Tahun 2016  
oleh  
Kementerian  
Kesehatan RI

### Faktor-faktor HIV/AIDS

1. Pergaulan bebas
2. Ketidakpedulian
3. Angka buta huruf
4. Kemiskinan
5. Penggunaan narkoba dan minuman beralkohol
6. Stigma terhadap HIV/AIDS
7. Faktor budaya
8. Kekurangan terhadap layanan kehamilan
9. Konflik suku dan perang saudara
10. Angka migrasi penduduk



### Faktor-faktor HIV/AIDS

1. **Pengguna kontrasepsi**
2. **Jumlah dokter terhadap populasi**
3. **Proporsi populasi masyarakat Muslim**
4. **Kehamilan saat remaja**
5. **Waktu belajar penduduk**



## Tinjauan Pustaka

### Faktor-faktor HIV/AIDS



Tingkat  
Kemiskinan

Prevalensi  
Penggunaan  
Narkoba



Tingkat  
Buta Huruf

Jumlah  
Lokalisasi



Ketersediaan  
Dokter



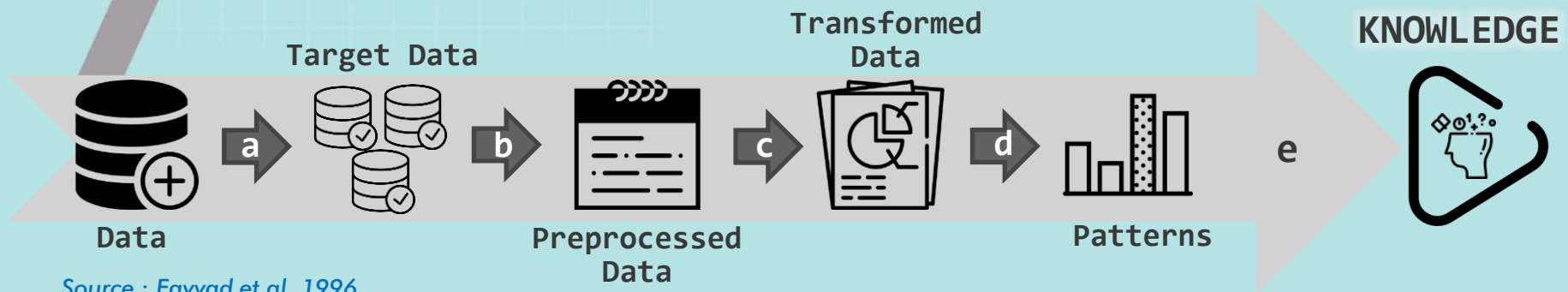


Tinjauan Pustaka

# da · ta min · ing



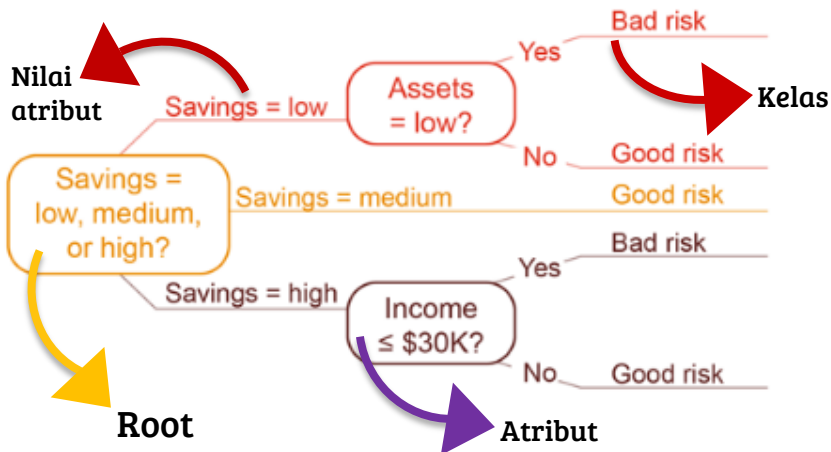
<u>Knowledge Discovery in Database</u>	<u>Tipe Data Mining</u>
a. Selection / data cleaning	❖ Klasterisasi
b. Proprecessing / data integrating	❖ Klasifikasi
c. Data transformation	❖ Regresi
d. Data mining	❖ Deteksi anomali
e. Interpretation / knowledge presentation	❖ Aturan Asosiasi
	❖ Perangkat lunak



## Tinjauan Pustaka

# Algoritma Pohon Keputusan

**Pohon Keputusan** : diagram alir yang membentuk struktur hirarki menyerupai struktur pohon, untuk mengambil sebuah keputusan secara bertahap dari ukuran atribut tertentu



Source : Larose, 2005

## Penentuan Ukuran Atribut

### Information Gain

$$InfGain(S, A) = Entropy(S) - \sum_v \frac{|S_v|}{|S|} [Entropy((S_v))]$$

$$Entropy(S) = - \sum_v p_j \log_2 p_j$$

### Gini Index

$$Gini Index(S) = 1 - \sum_v p_j^2$$

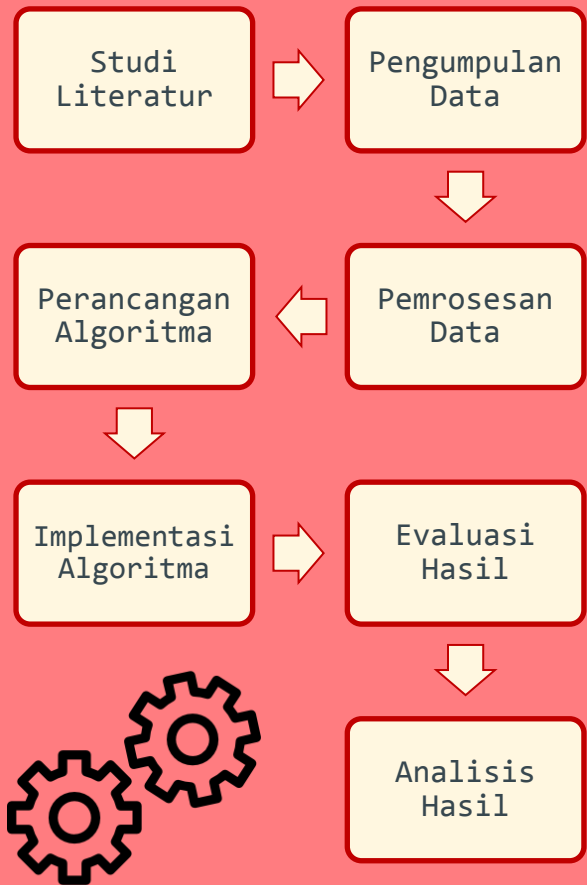
di mana:

$S$  : himpunan kasus

$k$  : banyaknya partisi  $S$

$p_j$  : rasio jumlah sampel di kelas  $j$  dengan jumlah semua sampel pada himpunan data

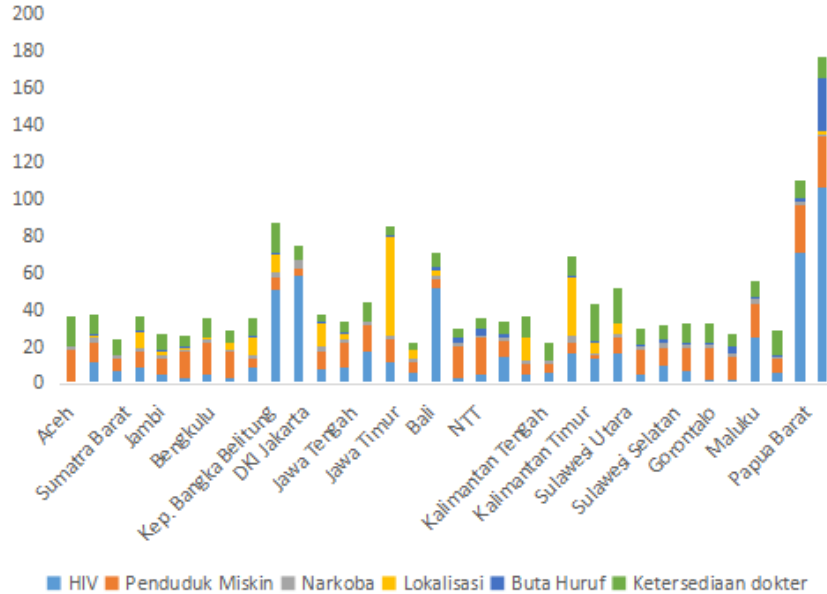
# Metodologi Penelitian



Variabel	Deskripsi	Sumber
Prevalensi HIV	Jumlah pengidap HIV tiap 100.000 penduduk pada tahun 2014	Kementerian Kesehatan
Persentase penduduk miskin	Jumlah penduduk miskin terhadap jumlah total penduduk pada tahun 2014	Badan Pusat Statistik
Prevalensi pengguna narkoba	Jumlah pengguna narkoba dan obat-obatan terlarang tiap 100.000 penduduk yang berusia 10-59 tahun pada tahun 2014	Badan Narkotika Nasional
Jumlah lokalisasi	Banyaknya lokalisasi prostitusi yang belum ditutup dalam provinsi tahun 2014	Kementerian Sosial
Persentase penduduk buta huruf	Jumlah penduduk yang buta huruf terhadap total jumlah penduduk yang berusia 14-44 tahun pada tahun 2014	Badan Pusat Statistik
Ketersediaan dokter	Jumlah dokter di Pusat Kesehatan Masyarakat yang tersedia tiap 100 orang penduduk	Badan Pusat Statistik

Deskripsi dan Sumber Data

# Hasil dan Analisis



Hasil Penentuan  
Ukuran Atribut

Information Gain  
Akurasi : 97,06%

Gini Index:  
Akurasi 91,17%

Variabel	N	Min	Maks	Mean	Median	SD
<b>Prevalensi HIV</b>	<b>34</b>	<b>1,222</b>	<b>106,049</b>	<b>16969</b>	<b>8,266</b>	<b>23,305</b>
Persentase penduduk miskin	34	4,09	27,8	11532	9,695	5,989
Prevalensi pengguna narkoba	34	1,21	5,01	2,032	1,865	0,704
Jumlah lokalisasi	34	0	12,51	3,15	2,505	2,515
Persentase penduduk buta huruf	34	0,08	28,5	1,87	0,61	4,814
Ketersediaan dokter	34	3,06	20,38	9,284	9,015	4,017

Prevalensi HIV

Prevalensi  
Rendah  
( $< 8,266$ )

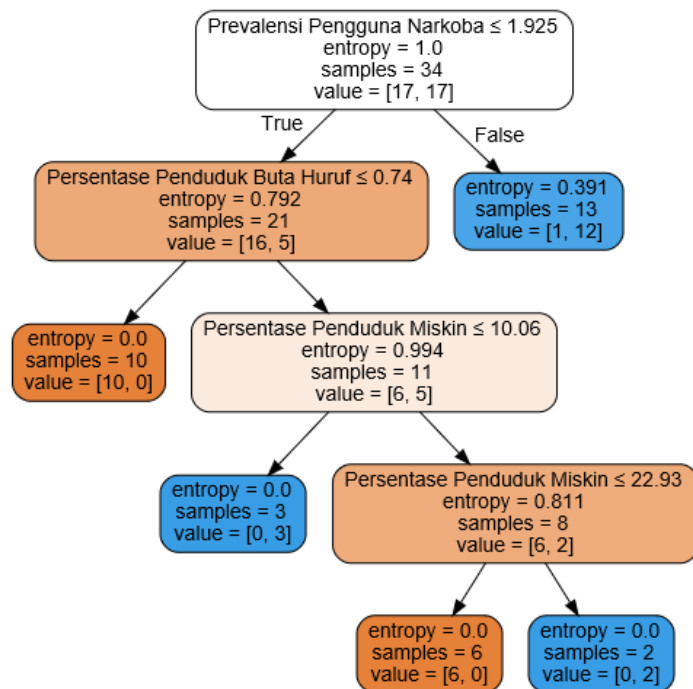
Prevalensi  
Tinggi  
( $> 8,266$ )

## Hasil dan Analisis

# Hasil Penentuan Ukuran Atribut

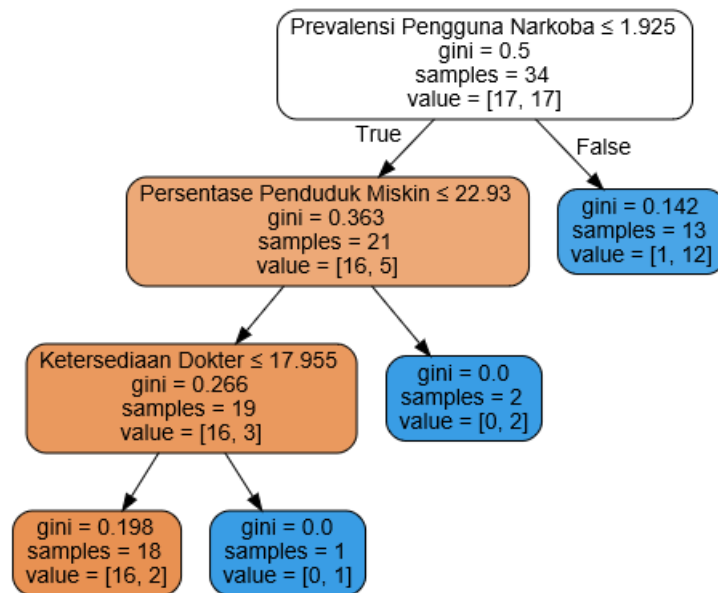
Hasil kriteria *Information Gain*

(akurasi: 97,06%)

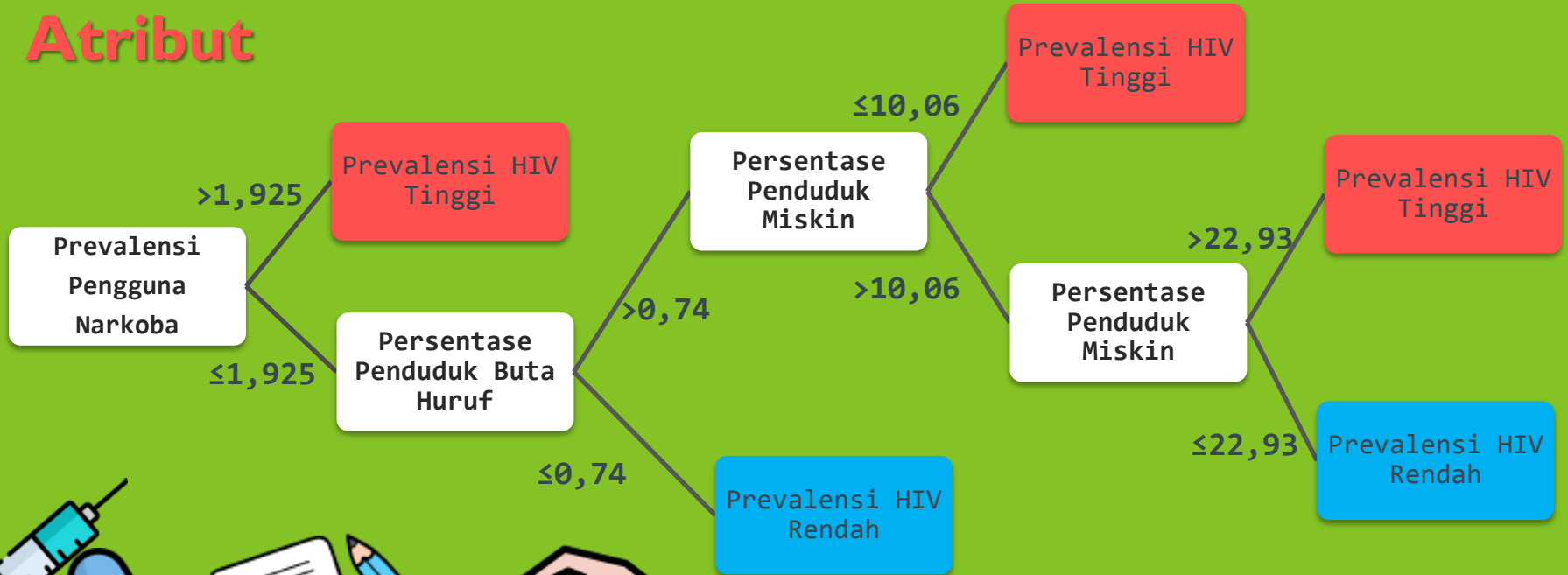


Hasil kriteria *Gini Index*

(akurasi: 91,17%)



# Hasil dan Analisis Hasil Penentuan Ukuran Atribut



# Hasil dan Analisis Hasil Penentuan Ukuran Atribut



Nilai Treshold = 1,925  
Nilai rata-rata atribut = 20,032

13 Provinsi  
Terkralifikasi  
Prevalensi HIV tinggi

Prevalensi pengguna narkoba tertinggi berada di Provinsi **DKI Jakarta** dengan nilai **5,01** memiliki prevalensi HIV bernilai **58,072** (tiap 100.000 penduduk), tertinggi kedua setelah Provinsi Papua.

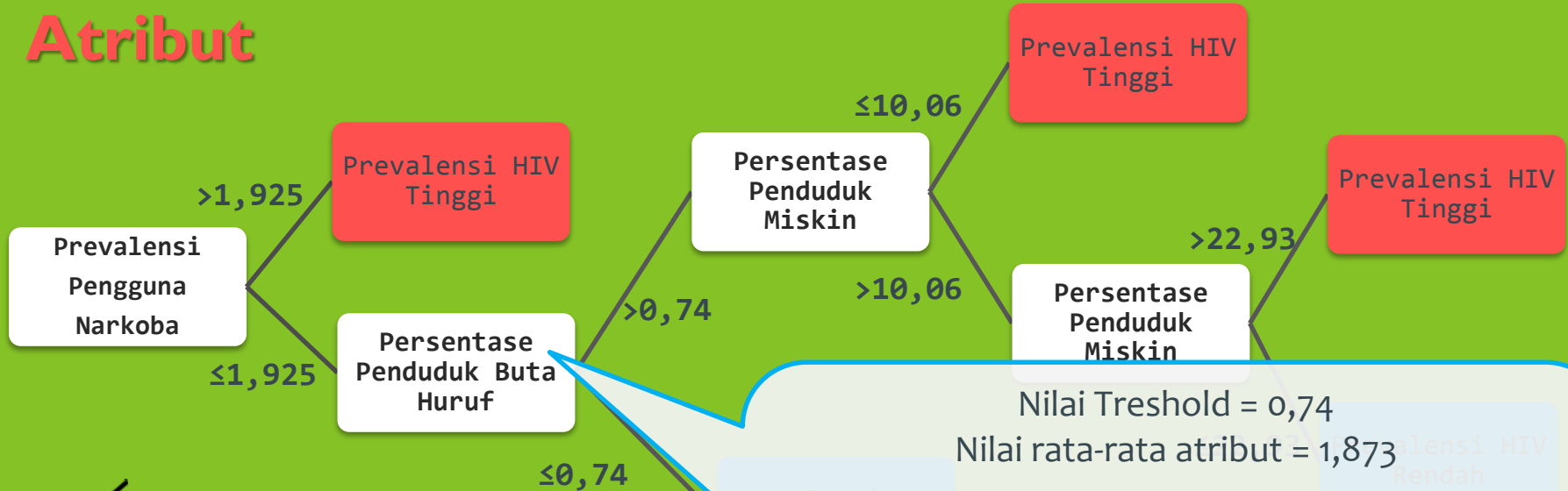
≤22,93 Prevalensi HIV Rendah

≤0,74

≤1,925

>1,925

# Hasil dan Analisis Hasil Penentuan Ukuran Atribut



Nilai Treshold = 0,74

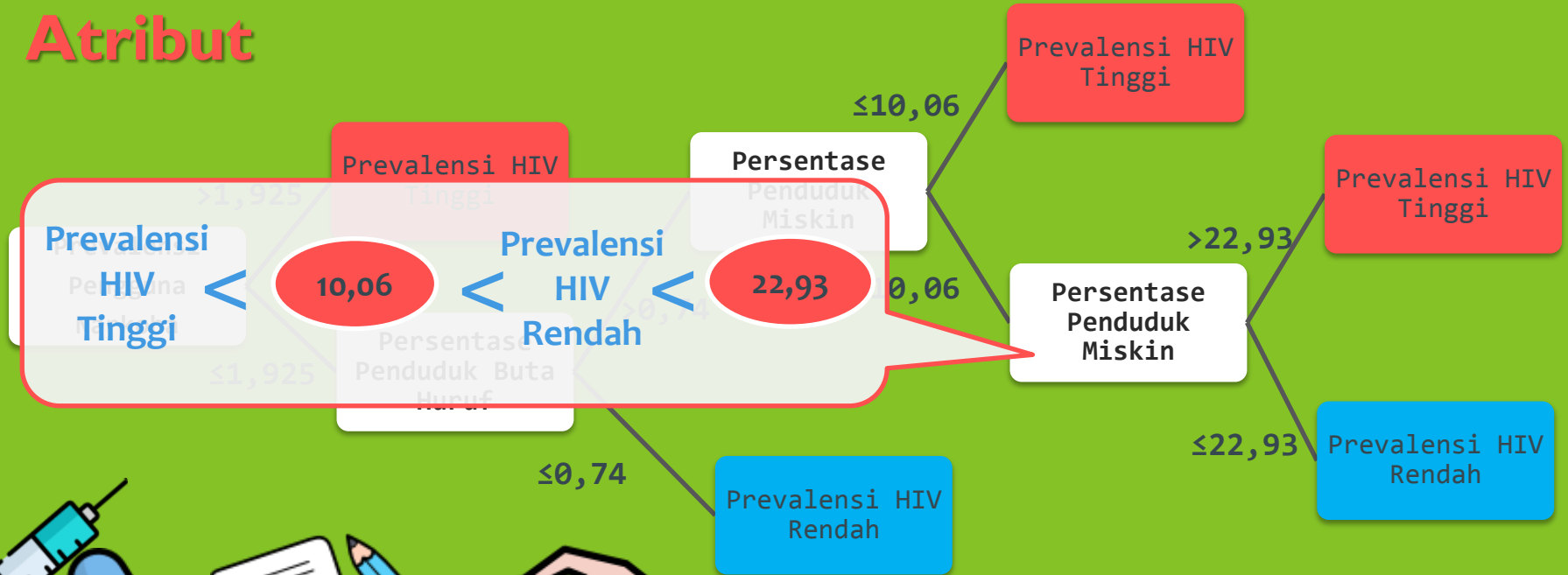
Nilai rata-rata atribut = 1,873

**Papua** merupakan provinsi dengan **penduduk buta huruf tertinggi** di Indonesia yakni bernilai **28,5%** sedangkan nilai **prevalensi HIV** pada angka **106,04** yang merupakan **tertinggi di Indonesia**





# Hasil dan Analisis Hasil Penentuan Ukuran Atribut



## Kesimpulan

1. Algoritma pohon keputusan dapat diimplementasikan menggunakan kriteria **Information Gain** dan **Gini Index**. Pada penelitian digunakan kriteria Information Gain karena menghasilkan akurasi yang lebih besar.
2. **Prevalensi pengguna narkoba, persentase penduduk buta huruf dan persentase penduduk miskin** adalah faktor yang paling berpengaruh terhadap penyebaran HIV/AIDS di Indonesia.
3. Analisis dari hasil menunjukkan secara umum **sesuai dengan hipotesis dari literatur, kecuali pada faktor Persentase Penduduk Miskin**

Hasil Penentuan  
Ukuran Atribut

Information Gain  
Akurasi : 97,06%

Gini Index:  
Akurasi 91,17%



## Saran

1. Untuk mengurangi angka prevalensi HIV/AIDS pemerintah lebih **fokus** dalam membenahi masalah **narkoba, buta huruf dan kemiskinan** di Indonesia.
2. Masyarakat Indonesia harus **lebih aktif** membantu pemerintah membenahi masalah tersebut di daerahnya masing-masing.
3. Algoritma pohon keputusan dapat diimplementasikan pada **kasus lain** dengan data-data pendukung untuk menyelesaikan masalah di Indonesia.
4. Untuk mendapatkan hasil yang lebih teliti pada penelitian selanjutnya dapat digunakan **skala yang lebih kecil**, misalnya skala kabupaten/kota sebagai objek data.





# *Terima Kasih*

*Semarang, 8 September 2018*