

# KLASIFIKASI TINGKAT KEMISKINAN BERDASARKAN PROVINSI DI INDONESIA MENGGUNAKAN ALGORITMA DECISION TREE

Athallah Kresna Bhadra<sup>1</sup>, Bryan Nicholas Marcos<sup>2</sup>, Rizky Bagus Setiawan<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika, Indonesia

## Informasi Artikel

Terbit: Januari 2026

## Kata Kunci:

Kemiskinan  
Decision Tree  
Klasifikasi  
Machine Learning  
Data Statistik

## ABSTRAK

Di Indonesia, isu kemiskinan masih menjadi fokus utama dalam pembangunan. Dengan memanfaatkan algoritma Decision Tree, penelitian ini berupaya untuk mengklasifikasikan tingkat kemiskinan di setiap provinsi di Indonesia. Data ini diambil dari Badan Pusat Statistik (BPS), yang mengindikasikan jumlah warga yang hidup dalam kemiskinan di setiap provinsi. Tingkat kemiskinan diklasifikasikan ke dalam tiga kategori; Rendah (di bawah 7%), Sedang (antara 7 dan 14%), serta Tinggi (lebih dari 14%). Proses pengolahan data dilaksanakan dengan menggunakan pustaka scikit-learn dan bahasa pemrograman Python. Tahapan yang dilakukan mencakup preprocessing, pembagian data menjadi data latih dan data uji (80:20), serta evaluasi model dengan memanfaatkan confusion matrix dan cross-validation. Hasil dari pengujian menunjukkan bahwa model Decision Tree mencapai akurasi 100%, dan semua data berhasil diklasifikasikan secara tepat. Penelitian ini mengungkapkan bahwa metode Decision Tree efektif dalam mengklasifikasikan tingkat kemiskinan antarprovinsi. Penelitian ini juga membuktikan bahwa model ini dapat dijadikan dasar untuk analisis data kebijakan sosial ekonomi.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Athallah Kresna Bhadra

Email: [athallahbhadra@gmail.com](mailto:athallahbhadra@gmail.com)

## 1. PENDAHULUAN

Kemiskinan adalah sebuah permasalahan ekonomi sosial yang masih dianggap krusial dalam upaya pembangunan nasional di Indonesia hingga saat ini. Badan Pusat Statistik (BPS) melaporkan bahwa pada tahun 2023, proporsi penduduk yang tergolong miskin di Indonesia mencapai 9,36%, meskipun data tersebut menunjukkan adanya penurunan jika dibandingkan dengan tahun-tahun sebelumnya [1]. Meskipun demikian, keadaan kemiskinan antarprovinsi menunjukkan perbedaan yang cukup besar, di mana beberapa daerah seperti Papua dan Nusa Tenggara Timur memiliki tingkat kemiskinan yang jauh lebih tinggi dibandingkan dengan provinsi lainnya seperti DKI Jakarta dan Bali [2].

Kemiskinan tidak hanya dapat dilihat sebagai kurangnya pendapatan, tetapi juga sebagai suatu fenomena yang mempunyai banyak dimensi, mencakup akses yang terbatas terhadap pendidikan, pelayanan kesehatan, dan peluang ekonomi [3]. Oleh karena itu, dalam menganalisis kemiskinan, diperlukan pendekatan yang dapat menangkap kompleksitas pola serta karakteristik data sosial ekonomi secara keseluruhan. Metode analisis yang umum dan bersifat deskriptif sering kali tidak memadai untuk mendeteksi pola hubungan antara variabel kemiskinan yang bersifat tidak linier dan memiliki banyak dimensi [4].

Perkembangan dalam bidang teknologi informasi serta kecerdasan buatan menyediakan kesempatan baru untuk menganalisis data sosial ekonomi dengan menggunakan machine learning. Machine learning memberikan kemampuan kepada sistem komputer untuk memahami pola dari data yang ada sebelumnya dan secara otomatis memproduksi prediksi atau klasifikasi tanpa adanya aturan yang telah ditentukan sebelumnya

[5]. Salah satu metode supervised learning yang sering dipakai untuk klasifikasi data adalah algoritma Decision Tree.

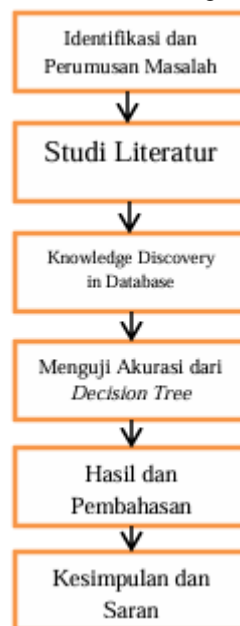
Algoritma Decision Tree menawarkan keuntungan dalam menyederhanakan masalah yang rumit menjadi bentuk pohon keputusan yang lebih mudah dimengerti dan divisualisasikan. Metode ini beroperasi dengan memisahkan data menurut atribut yang paling signifikan dengan memakai ukuran seperti *Gini Impurity* atau *Information Gain* [6]. Lebih dari itu, Decision Tree dianggap memiliki tingkat pemahaman yang sangat baik sehingga sangat cocok untuk menganalisis kebijakan publik yang didasarkan pada data [4].

Konsep klasifikasi dalam data mining adalah tindakan mengorganisir data ke dalam kelompok tertentu berdasarkan ciri-ciri yang terdapat padanya. Metode ini memanfaatkan data yang sudah diberi label sebagai data pembelajaran untuk menciptakan model prediksi bagi data yang belum dianalisis [7]. Klasifikasi sering diterapkan dalam analisis sosial dan ekonomi karena dapat memberikan gambaran terperinci tentang kondisi wilayah atau komunitas tertentu [2].

Penelitian sebelumnya telah menunjukkan bahwa Decision Tree cukup efektif untuk digunakan dalam mengklasifikasikan data sosial ekonomi. [8] menunjukkan bahwa algoritma ini memiliki kemampuan untuk mengkategorikan data kesejahteraan dengan tingkat akurasi yang lebih unggul jika dibandingkan dengan metode regresi linier. Temuan sejenis juga didapat oleh [7] yang mengungkapkan bahwa Decision Tree dapat menciptakan model klasifikasi yang tidak hanya akurat tetapi juga mudah untuk dipahami. Di samping Decision Tree, ada penelitian lainnya yang mengangkat metode klasifikasi yang tinggi dengan pendekatan yang berbeda, [9] menerapkan algoritma K-Nearest Neighbor dalam klasifikasi kedalaman kemiskinan dan meraih tingkat akurasi yang tinggi, meskipun terdapat kelemahan dalam hal interpretasi model. Oleh karena itu, Decision Tree dipilih untuk penelitian ini sebab mampu mengintegrasikan akurasi yang tinggi dengan kualitas interpretasi model yang baik dalam mengklasifikasikan tingkat kemiskinan provinsi di Indonesia berdasarkan data persentase penduduk miskin dari BPS.

## 2. METODE PENELITIAN

Pada penelitian untuk model Klasifikasi Tingkat Kemiskinan Berdasarkan Provinsi di Indonesia Menggunakan Algoritma Decision Tree. Gambar 1 adalah langkah dari penelitian yang dilakukan penulis.



**Gambar 1.** Langkah Penelitian

(Sumber : Data Penelitian, 2025)

Penjelasan berdasarkan langkah dari penelitian sesuai dengan model atau gambar diatas, yaitu:

- Tahap pertama melakukan penentuan identifikasi dan perumusan masalah pada penelitian ini.
- Langkah berikutnya adalah melakukan studi pustaka dengan membaca berbagai sumber untuk memahami penelitian yang sudah ada terutama membahas tentang cara mengklasifikasikan tingkat kemiskinan berdasarkan provinsi di Indonesia menggunakan algoritma Decision Tree.

- c. Langkah berikutnya adalah mengolah data dengan menggunakan tahapan Knowledge Discovery in Database atau KDD dari proses memilih data, membersihkan data, dan mengubah data sampai data tersebut siap untuk diolah dengan algoritma Decision Tree.
- d. Langkah berikutnya adalah menguji algoritma Decision Tree dengan cara melakukan percobaan membagi data menjadi dua bagian, yaitu data latih dan data uji, lalu melakukan percobaan untuk mendapatkan hasil akurasi yang paling tinggi.
- e. Hasil dan pembahasan adalah bagian yang menjelaskan temuan dari penelitian yang dilakukan oleh peneliti. Temuan ini menunjukkan seberapa akurat klasifikasi yang telah dilakukan mengenai data persentase tingkat kemiskinan di setiap provinsi di Indonesia.
- f. Kesimpulan ini adalah langkah di mana peneliti menentukan hasil akhir dari penelitian yang dijalankan oleh peneliti.

### 3. HASIL DAN ANALISIS

#### 3.1. Hasil Penelitian

Penelitian ini menghasilkan model klasifikasi tingkat kemiskinan berdasarkan data Persentase Penduduk Miskin per Provinsi di Indonesia, berikut adalah laman sumbernya: <https://opendata.jabarprov.go.id/id/dataset/persentase-penduduk-miskin-berdasarkan-provinsi-di-indonesia>. Data berikut bersumber dari publikasi resmi Badan Pusat Statistik (BPS) yang berisi persentase penduduk miskin, kode provinsi, nama provinsi, satuan, dan tahun.

Penelitian dimulai dengan membaca data menggunakan pustaka pandas, lalu dilakukan pembersihan data untuk memastikan tidak ada nilai yang hilang atau duplikat. Data dibagi menjadi tiga kelas berdasarkan ambang batas (threshold) yang telah ditetapkan adalah:

**Tabel 1.** Threshold

Kategori	Kriteria Persentase Kemiskinan
Rendah	< 7%
Sedang	7% - 14%
Tinggi	> 14%

Sumber : [10]

Proses klasifikasi ini menghasilkan variabel target (variabel Y) yang diberi nama kategori, yang berfungsi sebagai variabel dependen dalam model klasifikasi. Variabel independen mencakup Kode\_Provinsi, Persentase\_Penduduk\_Miskin, dan tahun.

Setelah data disiapkan, dataset dibagi menjadi 80% data latih dan 20% data uji dengan menggunakan fungsi `train_test_split`. Model Decision Tree yang telah dilatih dengan menggunakan kriteria Gini Impurity kemudian diuji pada data uji untuk mendapatkan nilai akurasi, precision, recall, dan F1-Score.

#### 3.2. Hasil Pengujian

Setelah data diklasifikasikan, algoritma Decision Tree diterapkan menggunakan bahasa pemrograman Python. Metode ini dipilih karena kemampuannya dalam menggambarkan pola hubungan antar variabel dan menciptakan model klasifikasi yang mudah dipahami dengan melihat struktur pohon keputusan.

##### 3.2.1. Proses Pengujian

Proses pengujian dilakukan untuk mengukur kemampuan model Decision Tree dalam mengklasifikasikan tingkat kemiskinan berdasarkan persentase kemiskinan di setiap provinsi. Tahapan pengujian mencakup:

- a. Melatih model menggunakan data latih ( $X_{train}$ ,  $Y_{train}$ ).
- b. Menguji model dengan memanfaatkan data uji ( $X_{test}$ ,  $Y_{test}$ ).
- c. Menentukan nilai akurasi dari model.
- d. Melakukan Cross-Validation untuk menilai kestabilan model.
- e. Mengevaluasi model melalui Confusion Matrix dan Laporan Klasifikasi.

##### 3.3.2. Hasil Evaluasi

Hasil evaluasi mengindikasikan bahwa model Decision Tree dapat mengklasifikasikan tingkat kemiskinan dengan akurasi yang sangat baik. Akurasi model dihitung memakai fungsi `accuracy_score` dari `scikit-learn`. Dari hasil pengujian model, diperoleh hasil sebagai berikut:

---

*Klasifikasi Tingkat Kemiskinan Berdasarkan Provinsi di Indonesia Menggunakan Algoritma Decision Tree.*  
(Athallah Kresna Bhadra)

```

Hasil Cross Validation (5-Fold):
Nilai Akurasi per Fold: [1. 1. 1. 1. 1.]
Rata-Rata Akurasi: 1.0

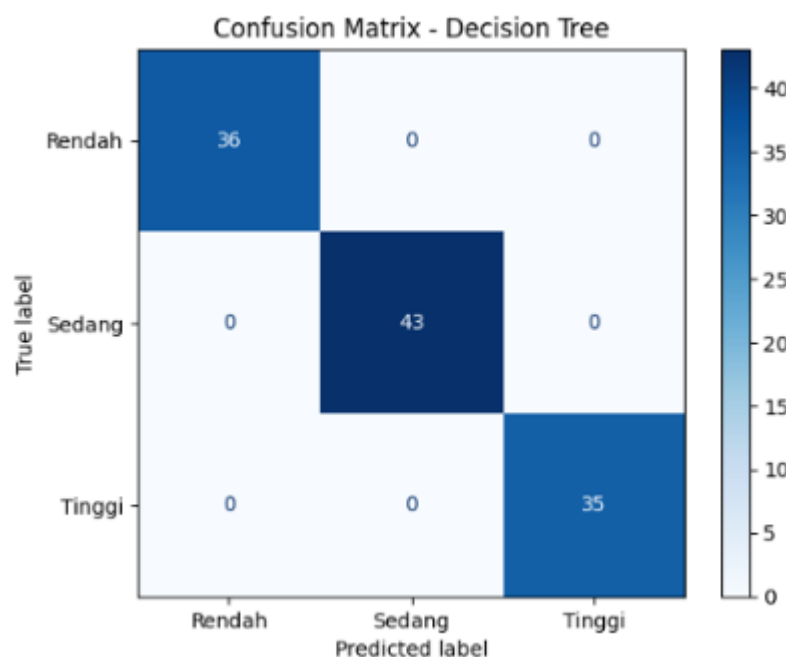
```

**Gambar 2.** Akurasi dan Cross\_Validation

Hasil pengujian yang dilakukan dengan metode 5-Fold Cross Validation menunjukkan bahwa model Decision Tree mencapai nilai akurasi sempurna, yaitu 1.0 atau 100% pada setiap fold. Ini menunjukkan bahwa dalam lima kali pembagian dan pengujian data secara silih berganti, model selalu berhasil memprediksi label kategori dengan tepat tanpa kesalahan.

### 3.3.3. Visualisasi Confusion Matrix

Confusion matrix berfungsi untuk memperlihatkan distribusi hasil klasifikasi antara kelas yang diprediksi dan kelas yang sebenarnya. Hasil dari confusion matrix ditunjukkan sebagai berikut:



**Gambar 3.** Confusion Matrix

Dari matriks ini, kita dapat melihat bahwa 36 data dalam kategori “Rendah” berhasil diidentifikasi dengan tepat sebagai “Rendah”, 43 data dalam kategori “Sedang” juga terprediksi dengan benar sebagai “Sedang”, 35 data kategori “Tinggi” diidentifikasi dengan akurat sebagai “Tinggi”, dan tidak ada satu pun data yang salah klasifikasinya atau berbeda dari diagonal yang bernilai nol. Oleh karena itu, model ini memberikan tingkat akurasi 100% atau klasifikasi sempurna, karena total prediksi yang benar sama dengan total data uji ( $36 + 43 + 35 = 114$ ).

Oleh karena itu, visualisasi confusion matrix di atas, dapat disimpulkan bahwa model Decision Tree memiliki kemampuan klasifikasi yang sangat baik terhadap tingkat kemiskinan dengan akurasi mencapai 100%. Tidak ada kesalahan dalam prediksi antar kelas yang menunjukkan bahwa model ini sangat tepat dan efisien. Kinerja model tetap konsisten dengan hasil evaluasi sebelumnya yang menunjukkan kepastian tinggi dari model ini. Temuan ini mengindikasikan bahwa persentase penduduk yang hidup dalam kemiskinan menjadi indikator tunggal yang sangat kuat dan efektif dalam mengklasifikasikan tingkat kemiskinan menjadi tiga kategori utama.

### 3.3.4. Laporan Klasifikasi

Evaluasi lebih lanjut dilaksanakan dengan memanfaatkan laporan klasifikasi, yang menyajikan metrik precision, recall, dan f1-score untuk setiap kategori Rendah, Sedang, dan Tinggi. Hasil dari evaluasi yang menggunakan laporan klasifikasi disajikan dalam Gambar 4.

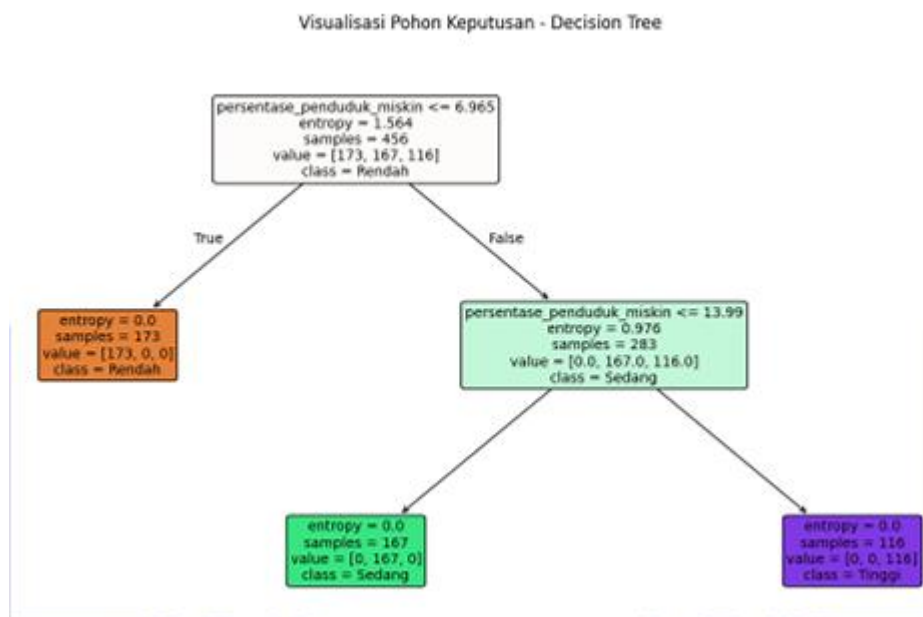
Laporan Klasifikasi:				
	precision	recall	f1-score	support
Rendah	1.00	1.00	1.00	36
Sedang	1.00	1.00	1.00	43
Tinggi	1.00	1.00	1.00	35
accuracy			1.00	114
macro avg	1.00	1.00	1.00	114
weighted avg	1.00	1.00	1.00	114

**Gambar 4.** Laporan Klasifikasi

Dari temuan tersebut, kita dapat menyimpulkan bahwa model ini menunjukkan kinerja yang seimbang di antara tiga kategori kemiskinan, dengan f1-score sebesar 1.00 atau 100%, yang menunjukkan adanya keseimbangan antara tingkat presisi dan sensitivitas model.

### 3.3.5. Visualisasi Pohon Keputusan

Visualisasi pohon keputusan menunjukkan struktur logika model untuk mengklasifikasikan tingkat kemiskinan. Hasil dari visualisasi pohon keputusan dapat dilihat pada Gambar 5.



**Gambar 5.** Visualisasi Pohon Keputusan

Visualisasi ini mengilustrasikan bahwa variabel *Persentase\_Penduduk\_Miskin* bertindak sebagai node utama dalam pembentukan pohon keputusan. Cabang-cabang berikutnya terbentuk berdasarkan nilai ambang batas atau threshold, yakni 7% dan 14% yang sesuai dengan kategori kemiskinan.

Dengan demikian, temuan ini menunjukkan bahwa variabel *Persentase\_Penduduk\_Miskin* menjadi elemen kunci dalam menentukan klasifikasi tingkat kemiskinan antarprovinsi di Indonesia. Bentuk pohon yang sederhana dan mudah dipahami menunjukkan keuntungan dari Decision Tree dalam menyajikan interpretasi hasil secara jelas.

#### 4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Decision Tree dapat dengan tepat dan jelas mengklasifikasikan tingkat kemiskinan antarprovinsi di Indonesia. Model ini menunjukkan tingkat akurasi mencapai 100% dan dapat menjelaskan dengan baik batasan kategori kemiskinan. Untuk studi yang akan datang, disarankan untuk memasukkan variabel sosial ekonomi tambahan seperti tingkat pendidikan, angka pengangguran, dan PDRB agar model dapat menawarkan analisis yang lebih menyeluruh.

#### DAFTAR PUSTAKA

- [1] BPS, “Statistik Indonesia 2024,” *Stat. Indones.* 2024, vol. 1101001, p. 790, 2024, Accessed: Nov. 01, 2025. [Online]. Available: <https://www.bps.go.id/id/publication/2024/02/28/c1bacde03256343b2bf769b0/statistik-indonesia-2024.html>
- [2] N. P. N. Hendayanti and M. Nurhidayati, “KLASIFIKASI TINGKAT KEPARAHAN KEMISKINAN PROVINSI DI INDONESIA DENGAN ANALISIS DISKRIMINAN,” *Math Educ. J.*, vol. 5, no. 1, pp. 14–21, May 2021, doi: 10.15548/MEJ.V5I1.2510.
- [3] W. Bank, “Poverty and Shared Prosperity 2022 : Correcting Course,” 2022, Accessed: Nov. 13, 2025. [Online]. Available: [https://www.google.co.id/books/edition/Poverty\\_and\\_Shared\\_Prosperty\\_2022/Z9GgEAAAQBAJ?hl=id&gbpv=1&dq=World+Bank.+\(2022\).+Poverty+and+Shared+Prosperity+2022:+Correcting+Course.&pg=PA1950&printsec=frontcover](https://www.google.co.id/books/edition/Poverty_and_Shared_Prosperty_2022/Z9GgEAAAQBAJ?hl=id&gbpv=1&dq=World+Bank.+(2022).+Poverty+and+Shared+Prosperity+2022:+Correcting+Course.&pg=PA1950&printsec=frontcover)
- [4] K. P. . Murphy, “Probabilistic machine learning : an introduction,” p. 826, 2022, Accessed: Nov. 12, 2025. [Online]. Available: [https://www.google.co.id/books/edition/Probabilistic\\_Machine\\_Learning/OyYuEAAAQBAJ?hl=id&gbpv=1&dq=Murphy,+K.+P.+\(2022\).+Probabilistic+Machine+Learning:+An+Introduction.+MIT+Press.&pg=PA814&printsec=frontcover](https://www.google.co.id/books/edition/Probabilistic_Machine_Learning/OyYuEAAAQBAJ?hl=id&gbpv=1&dq=Murphy,+K.+P.+(2022).+Probabilistic+Machine+Learning:+An+Introduction.+MIT+Press.&pg=PA814&printsec=frontcover)
- [5] S. K. M. K. Rahmaddeni, S. K. M. K. Denok Wulandari, M. Renova, A. M. Gilang Ramadhan, and R. Sari, *Machine Learning*. Serasi Media Teknologi, 2024. Accessed: Nov. 20, 2025. [Online]. Available: <https://books.google.co.id/books?id=owoOEQAQBAJ>
- [6] D. Setiawati, I. Taufik, W. Z. Budiawan, J. Teknik Informatika, and F. Sains dan Teknologi, “Klasifikasi Terjemahan Ayat Al-Quran Tentang Ilmu Sains Menggunakan Algoritma Decision Tree Berbasis Mobile,” *J. Online Inform.*, vol. 1, no. 1, pp. 24–27, Jun. 2020, doi: 10.15575/JOIN.V1I1.7.
- [7] F. M. Hana, “Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, Oct. 2020, doi: 10.47970/SISKOM-KB.V4I1.173.
- [8] A. Ermillian and K. Nugroho, “Perancangan Model Deteksi Potensi Siswa Putus Sekolah Menggunakan Metode Logistic Regression Dan Decision Tree,” *J. Inform. J. Pengemb. IT*, vol. 9, no. 3, pp. 281–295, Dec. 2024, doi: 10.30591/JPIT.V9I3.8007.
- [9] R. C. Adelina, “Implementasi Algoritma K-Neraest Neighbor dalam penentuan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur,” 2022, Accessed: Nov. 13, 2025. [Online]. Available: <http://theses.uin-malang.ac.id/38201/>
- [10] “Persentase Penduduk Miskin Berdasarkan Provinsi di Indonesia.” Accessed: Nov. 13, 2025. [Online]. Available: <https://opendata.jabarprov.go.id/id/dataset/persentase-penduduk-miskin-berdasarkan-provinsi-di-indonesia>