

## Klasifikasi Dan Prediksi Keluarga Berisiko Stunting Di Provinsi Jambi Menggunakan Metode Knn Dan Naive Bayes

Niko Akbar <sup>1)\*</sup>, Hamzah Alghifari<sup>2)</sup>, Nurul Abdillah<sup>3)</sup>, Oki Dahwanu<sup>4)</sup>

<sup>1),2),3),4)</sup> Jurusan Sistem Informasi, Universitas Jambi

\*Corresponding Email : nikoakbar@unja.ac.id

---

### Abstrak

Di Indonesia Prevalensi stunting sebesar 37,2%, naik dari 35,6% pada tahun 2019 dan 36,8%, dengan mayoritas dipengaruhi oleh penduduk setempat. Kementerian Kesehatan Indonesia memperkirakan bahwa prevalensi stunting akan mencapai 38,9% pada tahun 2020. Permasalahannya Beberapa Data yang diambil dan dipakai berupa data sekunder yang di ada diwebsite opendata provinsi jambi yang berjudul Faktor Penapisan Keluarga Berisiko Stunting di Provinsi Jambi. Dengan menggunakan algoritma KNN dan Naive Bayes, maka didapatkan hasilnya berupa kedua algoritma cocok untuk pengklasteran dan Scoring dari algoritma menunjukkan hasil yang berbeda. Karena keakurasiannya sesuai dengan perhitungan manual yang telah dijabarkan pada bagian pengolahan data. Beberapa hasil dari cross-validasi menyatakan bahwa nilai accuracy 85,29%, nilai precision berupa 83,33%, nilai recall 85,29%.

**Kata Kunci:** Klasifikasi, Prediksi, KNN, Naive Bayes.

### Abstract

*In Indonesia, the prevalence of stunting is 37.2%, up from 35.6% in 2019 and 36.8%, with the majority affected by the local population. The Indonesian Ministry of Health estimates that the prevalence of stunting will reach 38.9% in 2020. The problem is that some of the data collected and used is secondary data found on the Jambi province open data website entitled Factors Screening Families at Risk of Stunting in Jambi Province. Using the KNN and Naive Bayes algorithms, the results show that both algorithms are suitable for clustering, and the scoring of the algorithms shows different results. This is because the accuracy is in line with the manual calculations described in the data processing section. Several results from cross-validation indicate an accuracy value of 85.29%, a precision value of 83.33%, and a recall value of 85.29%.*

**Keywords:** Classification, Prediction, KNN, Naive Bayes.

---

## PENDAHULUAN

Sebagian besar penduduk lokal menderita stunting, yang meningkat dari 35,6% pada tahun 2019 menjadi 37,2%. Pada tahun 2020, Kementerian Kesehatan

Indonesia memperkirakan 38,9% penduduk akan mengalami stunting (Indraputra, 2020). Rata-rata tinggi badan balita lebih tinggi atau lebih rendah daripada balita yang mengalami stunting. Menurut Organisasi Kesehatan Dunia (WHO), jika seseorang memiliki tinggi badan atau pendek yang tidak normal, tinggi badan atau panjang badannya dapat ditentukan menggunakan kriteria pertumbuhan umum untuk anak-anak[1].

Provinsi Jambi menargetkan angka stunting pada tahun 2024 sebesar 12 persen. Hal itu tidak lepas dari upaya pencegahan dan penanggulangan stunting yang semakin gencar dilakukan oleh pemerintah Bersama stakeholder terkait.

Data mining adalah Teknik yang digunakan untuk mengekstraksi atau “menambang” informasi dari kumpulan data yang besar. Data mining adalah meninjau berbagai informasi guna mendapat korelasi yang tak terduga dan merangkum informasi melalui metode yang berbeda agar bisa dimengerti dan berguna untuk pemilik informasi[2].

Permasalahan yang terjadi adalah bagaimana cara membangun model klasifikasi kabupaten berdasarkan produksi perkebunan rakyat tahunan, seberapa akurasi metode KNN dan Naive Bayes dalam klasifikasi ini, dan kabupaten mana yang termasuk dalam kategori penelitian penulis.

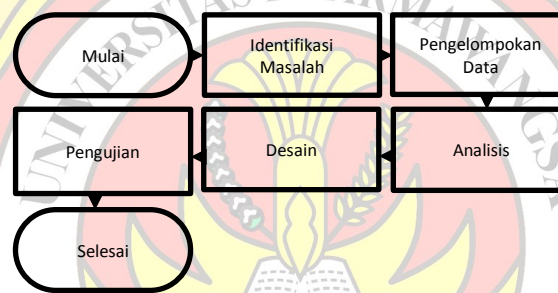
Sehingga Tujuan penelitian ini ditujukan untuk membangun model klasifikasi Stunting di Provinsi Jambi dan. membandingkan Performa Metode KNN dan Metode Naive Bayes, dan memberikan insight kebijakan berbasis klasifikasi tersebut.

Beberapa penelitian yang pernah dilakukan berupa Algoritma K-Nearest Neighbors telah digunakan dalam klasifikasi status gizi balita, ibu hamil, dan klasifikasi obesitas. Metode Modified K-Nearest Neighbor (M-KNN) telah digunakan dalam mengidentifikasi faktor risiko status gizi balita dengan akurasi sebesar 75% menggunakan data status gizi balita pada nilai  $k=4$ [3]. Kemudian ada penelitian lain menggunakan 412 data gizi balita, membandingkan metode Naive Bayes dan KNN

untuk klasifikasi status gizi balita diperoleh hasil akurasi 80,60% dengan Naive Bayes dan 91,79% dengan K-Nearest Neighbor[4].

Algoritma KNN telah digunakan untuk mengklasifikasikan sentimen publik mengenai vaksinasi COVID-19 menjadi kategori positif dan Negatif. Dengan nilai  $k=3$  diperoleh akurasi 85%[5] Kemudian terdapat penelitian berjudul Sistem Deteksi Pelat Kendaraan dengan menggunakan Metode K-Nearest Neighbour (KNN), nilai K yang optimal adalah 1 dengan akurasi 88,88% dalam jarak 40 cm.

## METODE PENELITIAN



Gambar 1. Alur Penelitian

### a. Bahan dan Alat Utama, Tempat

Penelitian ini menggunakan metode penelitian kuantitatif. Metode penelitian ini menggunakan data dalam bentuk angka dan analisis data. Data yang diambil berupa data berupa tahun 2019 - 2023.

penelitian ini dilakukan di wilayah provinsi jambi pada Badan Pusat statistik provinsi Jambi (<https://jambi.bps.go.id/>) dan Jambi Data & Analytic Center (<https://opendata.jambiprov.go.id/>). Penelitian ini dilakukan pada bulan september 2025 sampai dengan selesai

Data Sekunder dalam penelitian ini adalah data yang diperoleh melalui situs resmi [jambi.bps.go.id](http://jambi.bps.go.id) dan <https://opendata.jambiprov.go.id/> pada tahun 2019 – 2023

Data yang akan dipakai berupa data sekunder yang di open data provinsi jambi yang berjudul Faktor Penapisan keluarga berisiko stunting berdasarkan kecamatan di provinsi Jambi Tahun 2021 - 2024

Tabel 1 Daftar Faktor Penapisan Keluarga Berisiko Stunting

nama_kabupaten_kota	kode_kecamatan	nama_kecamatan	tahun	jumlah_keluarga_b erisiko_stunting	keluarga_tidak_memiliki_sumber_air_minum_layak	keluarga_tidak_memiliki_j amban_layak	pus_terlalu_muda_tidak_menggunakan_kb	pus_terlalu_tua_tidak_menggunakan_kb	pus_terlalu_dekat_tidak_menggunakan_kb
KERINCI	150105	AIR HANGAT	2021	874	25	217	7	473	28
KERINCI	150121	AIR HANGAT BARAT	2021	690	2	65	13	380	27
KERINCI	150111	AIR HANGAT TIMUR	2021	1.889	752	490	40	843	8
KERINCI	150107	BATANG MERANGIN	2021	1.283	124	464	24	580	37
KERINCI	150120	BUKITKERMAN	2021	1.245	167	314	29	573	34
KERINCI	150102	DANAU KERINCI	2021	1.260	27	199	14	646	57
KERINCI	150123	DANAU KERINCI BARAT	2021	1.037	113	307	7	464	38
KERINCI	150117	DEPATI TUJUH	2021	1.259	8	155	18	666	27
KERINCI	150106	GUNUNG KERINCI	2021	1.892	283	1.229	68	574	20
KERINCI	150101	GUNUNG RAYA	2021	795	209	195	11	331	55
KERINCI	150115	GUNUNG TUJUH	2021	1.610	60	405	59	718	23
KERINCI	150109	KAYU ARO	2021	2.404	211	922	78	1.114	19
KERINCI	150119	KAYU ARO BARAT	2021	1.786	200	219	29	966	44
KERINCI	150100	KELING DAJAU	2021	1.390	139	838	11	567	9
KERINCI	150104	SITINJAU LAUT	2021	782	32	245	6	393	19
KERINCI	150116	SULAK	2021	2.582	112	1.187	87	1.087	67
KERINCI	150110	SULAK MUKAI	2021	1.138	106	552	37	438	10
KERINCI	150122	TANAH COGOK	2021	786	10	302	8	368	17
MERANGIN	150202	BANGKO	2021	3.956	380	192	37	2.029	49
MERANGIN	150210	BANGKO BARAT	2021	1.170	332	178	16	532	15
MERANGIN	150212	BATANG MASUMAI	2021	1.143	118	285	19	460	10
MERANGIN	150201	JANGKAT	2021	1.528	850	408	64	431	6
MERANGIN	150218	JANGKAT TIMUR	2021	1.706	868	792	71	491	3
MERANGIN	150209	LEBAH MASURAI	2021	2.071	1.186	737	81	615	5
MERANGIN	150221	MARSO TABIR	2021	1.711	388	848	98	722	8
MERANGIN	150203	MUARA SIAU	2021	1.565	709	640	65	462	12
MERANGIN	150211	HALO TANTAN	2021	2.042	361	884	58	811	22
MERANGIN	150206	PAMENANG	2021	3.517	846	709	58	1.588	24
MERANGIN	150213	PAMENANG BARAT	2021	2.134	522	522	62	969	16
MERANGIN	150220	PAMENANG SELATAN	2021	1.118	170	244	33	504	11
MERANGIN	150217	PANGKALAN JAMBU	2021	832	88	296	27	313	2
MERANGIN	150219	RENAH PAIRILANG	2021	1.412	193	112	24	770	8
MERANGIN	150216	RENAH PEMBARAP	2021	1.585	264	397	74	508	16
MERANGIN	150204	SUNGAI MANAU	2021	1.572	520	556	48	462	14
MERANGIN	150205	TABIR	2021	3.831	705	2.236	165	1.253	28
MERANGIN	150223	TABIR BARAT	2021	1.728	1.159	1.287	27	429	9
MERANGIN	150214	TABIR ILIR	2021	1.281	159	696	55	530	2
MERANGIN	150222	TABIR LINTAS	2021	931	112	337	19	374	6
MERANGIN	150208	TABIR SELATAN	2021	2.796	501	340	38	1.390	30
MERANGIN	150215	TABIR TIMUR	2021	995	94	389	51	429	8

b. Teknik Analisis

Langkah – langkah dalam data mining adalah

a. Pembersihan data (Data Cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan.

b. Integrasi Data (Data Integration)

Integrasi Data merupakan penggabungan data dari berbagai database kedalam satu database baru.

c. Seleksi Data (Data Selection)

Data yang ada pada database seringkali tidak semuanya dipakai, oleh karena itu hanya ada data yang sesuai untuk dianalisis yang akan diambil dari database.

d. Transformasi Data (Data Transformation)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam Data Mining.

e. Proses Mining

Suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan Data Mining.

f. Evaluasi Pola (Pattern Evaluation)

Untuk mengidentifikasi pola-pola menarik ke dalam knowledge based yang ditemukan.

g. Presentasi Pengetahuan (Knowledge Presentation)[6].

Visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Normalisasi data

Normalisasi data merupakan proses mengubah skala data numerik ke dalam rentang yang seragam. Tujuan utamanya adalah untuk memastikan bahwa setiap fitur (variabel) memiliki pengaruh yang seimbang terhadap model. Sangat penting untuk algoritma yang sensitif terhadap skala. Normalisasi data yang dipakai dalam penelitian ini yaitu dengan melakukan teknik min-max normalization dengan persamaan[3]

$$\text{normalized}(x) = \frac{\text{range } x + (x \text{ minValue})(\text{maxRange} - \text{minRange})}{\text{maxValue} - \text{minValue}}$$

Rumus Persamaan Normalisasi[3]

Untuk memecahkan klasifikasi ini, digunakan 2 metode pengujian, yaitu persentase bersplit dan K-Fold cross validation. Pada persen split memiliki 8 : 2

yang artinya, data tersebut akan dipecah menjadi 2 bagian, yakni data training dan data testing. 80 persen dari data akan digunakan sebagai data pelatihan, dan 20 persen akan digunakan sebagai data evaluasi. Menggunakan K=10 untuk K-Fold Cross Validation.

Algoritma K-Nearest Neighbor (KNN) bekerja berdasarkan prinsip bahwa objek serupa berada dekat satu sama lain. Ia mengklasifikasikan data baru berdasarkan mayoritas kelas dari K tetangga terdekat, Sedangkan Naive Bayes merupakan algoritma probabilistik yang didasarkan oleh teorema Bayes. Segala atribut diperlukan secara bebas dan sama antara satu atribut dengan atribut yang lainnya. Metode ini memakai Naive Bayes Classifier dalam perhitungan nilai bobot peluang pada setiap atribut. Langkah-langkah untuk menghitung Metode Algoritma K-Nearest Neighbor :

- Menentukan Parameter K (Jumlah tetangga terdekat)
- Menghitung kuadrat jarak Euclid (query Instance) masing-masing objek terhadap data sampel yang diberikan.
- Mengurutkan objek tersebut ke dalam kelompok yang mempunyai jarak Euclid terkecil.
- Mengumpulkan kategori Y (Klasifikasi Nearest Neighbor)
- Dengan kategori Nearest Neighbor yang mayoritas, maka dapat diprediksi nilai query instance yang telah dihitung.
- Untuk mendefinisikan jarak antar dua titik yaitu titik pada data training (x) dan titik pada data testing (y) maka digunakan semua rumus Euclidean

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan :

d = Jarak Kedekatan

x = Data Training

y = Data Testing

n = Jumlah Atribut individu antara 1 sampai dengan n

$i$  = Atribut individu antara 1 sampai dengan  $n$ [7]

Algoritma Naive Bayes

Algoritma yang efisien dan efektif untuk pembelajaran mesin dan data mining. Berbasis Naive Bayes Classifier pada kombinasi Bayes Theorem dan Atribut Independensi anggapan. Naive Bayes Classifier didasarkan pada asumsi yang disederhanakan nilai-nilai atribut independen secara kondisional, dengan asumsi dari nilai target yang diberikan.

Klasifikasi

Terdapat banyak metode untuk klasifikasi dalam data mining seperti Naive Bayes dan KNN. Naive Bayes mempunyai kelebihan, yaitu dapat mencapai nilai akurasi yang tinggi dengan keterbatasannya jumlah. Sedangkan, Metode K-Nearest Neighbor mempunyai kelebihan mampu bekerja dengan maksimal terhadap data yang mempunyai kesalahan acak (noise)[4].

Algoritma klasifikasi data mining digunakan untuk mengukur akurasi algoritma klasifikasi, metode yang dapat digunakan antara lain cross validation, confusion matrix, dan kurva ROC (Receiver Operating Characteristics. Untuk mengembangkan aplikasi (development) berdasarkan model yang dibuat, digunakan Rapid Miner.

a. Cross Validation

Validation adalah pengujian standar yang dilakukan untuk memprediksi error rate. Data training dibagi secara random ke dalam beberapa bagian dengan perbandingan yang sama kemudian error rate dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh error rate untuk mendapatkan error rate secara keseluruhan.

b. Confusion Matrix

Metode yang menggunakan tabel matrix seperti tabel 2, Jika dataset hanya terdiri dari dua kelas, kelas yang satu dianggap positif dan yang lainnya negatif.

Tabel 2. Model Confusion Matrix

Classification		Prediction	
		Class = YES	Class = NO
Observed Class	Class = YES	A (True Positive TP)	B (False Negative FN)
	Class = NO	C (False Positive FP)	D (True Negative TN)

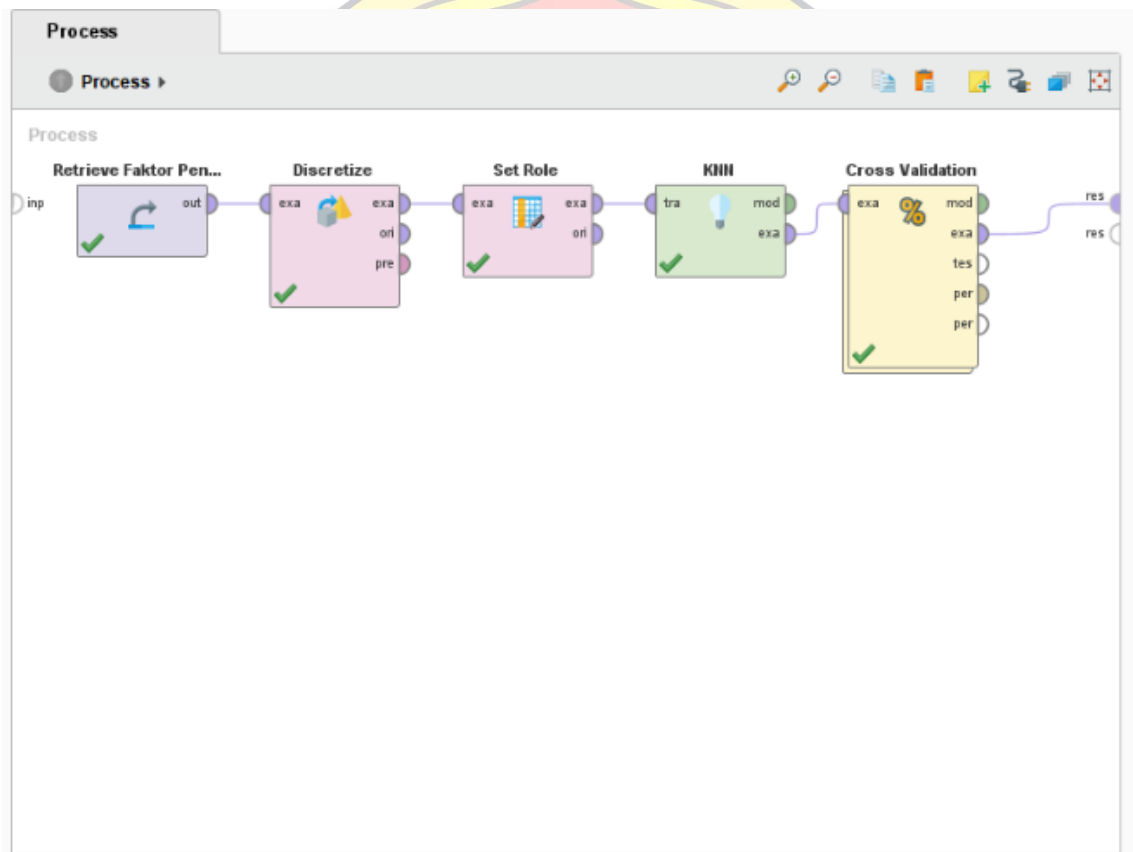
Atribut yang digunakan dalam setelah dibersihkan adalah

- No
- Nama Kabupaten Kota
- Nama Kecamatan
- Tahun
- Jumlah Keluarga berisiko stunting
- Jumlah Keluarga yang tidak memiliki sumber air minum layak
- Jumlah Keluarga tidak memiliki jamban layak
- Jumlah Pasangan Usia Subur (PUS) Muda ingin punya anak
- Jumlah Pasangan Usia Subur (PUS) Tua ingin punya anak
- Jumlah Pasangan Usia Subur (PUS) Usia berdekatan ingin punya anak
- Jumlah Pasangan Usia Subur (PUS) Usia berjauhan ingin punya anak

## HASIL DAN PEMBAHASAN

Algoritma K-Nearest Neighbors telah digunakan dalam klasifikasi status gizi balita, ibu hamil, dan klasifikasi obesitas. Metode Modified K-Nearest Neighbor (M-KNN) telah digunakan dalam mengidentifikasi faktor risiko status gizi balita dengan

akurasi sebesar 75% menggunakan data status gizi balita pada nilai  $k=4$ . Kemudian ada penelitian lain menggunakan 412 data gizi balita, membandingkan metode Naive Bayes dan KNN untuk klasifikasi status gizi balita diperoleh hasil akurasi 80,60% dengan Naive Bayes dan 91,79% dengan K-Nearest Neighbor. Algoritma KNN telah digunakan untuk mengklasifikasikan sentimen publik mengenai vaksinasi COVID-19 menjadi kategori positif dan Negatif. Dengan nilai  $k=3$  diperoleh akurasi 85%. Kemudian terdapat penelitian berjudul Sistem Deteksi Pelat Kendaraan dengan menggunakan Metode K-Nearest Neighbour (KNN), nilai K yang optimal adalah 1 dengan akurasi 88,88% dalam jarak 40 cm.



Gambar 2. Perancangan Data Mining dengan Metode K-Means dan KNN.

Dalam tahap akhir sebuah penelitian dan pengimplementasian sebuah sistem pengelompokan Keluarga berpotensi Stunting di Provinsi Jambi dengan menggunakan metode KNN dan K-Means telah mendapatkan hasil signifikan.

Beberapa hasil dari cross-validasi menyatakan bahwa nilai accuracy 85,29%, nilai precision berupa 83,33%, nilai recall 85,29%.

## SIMPULAN

Hal ini menjadikan kedua algoritma cocok untuk pengklasteran dan Scoring dari algoritma menunjukkan hasil yang berbeda. Karena keakurasiannya sesuai dengan perhitungan manual yang telah dijabarkan pada bagian pengolahan data.

## DAFTAR PUSTAKA

- [1] R. A. Indraputra and R. Fitriana, "K-Means clustering data COVID-19," *Jurnal Teknik Industri*, vol. 10, no. 3, pp. 275–282, 2020.
- [2] R. S. Cumel David Zamri, "Perbandingan Metode Data Mining untuk Prediksi Banjir dengan Algoritma Naïve Bayes dan KNN," *SENTIMAS: Seminar Nasional Penelitian dan ...*, pp. 40–48, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas/article/view/353%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/353/132>
- [3] M. Musnaimah, A. Alifatin, and N. Hayatin, "Expert System to Identify Risk Factors of Toddler's Nutrition Status with Case Based Reasoning," *Jurnal Perempuan dan Anak*, vol. 3, no. 1, p. 27, 2020, doi: 10.22219/jpa.v3i1.11810.
- [4] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web," *Jurnal Media Informatika Budidarma*, vol. 6, no. 2, p. 777, 2022, doi: 10.30865/mib.v6i2.3566.
- [5] S. Suprayogi, C. A. Sari, and E. H. Rachmawanto, "Sentiment Analyst on Twitter Using the K-Nearest Neighbors (KNN) Algorithm Against Covid-19 Vaccination," *Journal of Applied Intelligent System*, vol. 7, no. 2, pp. 135–145, 2022, doi: 10.33633/jais.v7i2.6734.
- [6] A. F. Watratan and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *Journal of Applied Computer Science and Technology*, vol. 1, no. 1, pp. 7–14, 2020.
- [7] S. T. M. K. Yahya, *Data Mining*. CV Jejak (Jejak Publisher), 2022. [Online]. Available: <https://books.google.co.id/books?id=0J2mEAAAQBAJ>