

## **Klasifikasi Curah Hujan di Provinsi Bali Berdasarkan Metode Naïve Bayesian**

**I Gede Aris Gunadi , Ayu Aprilyana Kusuma Dewi**  
**Jurusan Pendidikan Fisika FMIPA Undiksha , Program Magister Ilmu**  
**Komputer Pascasarjana Undiksha**  
**igedearisgunadi@undiksha.ac.id , aprilyana.beklencipss15@gmail.com**

### **Abstrak**

Wilayah Indonesia merupakan bagian wilayah tropis dengan intensitas curah hujan yang tinggi. Pada beberapa wilayah di Indonesia khususnya wilayah Bali sering muncul suatu fenomena alam yaitu bila saat musim hujan tiba terjadi limpahan air yang cukup banyak, berdasarkan data tahunan dari BMKG (Badan Meteorologi Klimatologi dan Geofisika) data curah hujan yang digunakan perlu dilakukan klasifikasi agar dapat memberikan informasi tentang pola curah hujan disuatu wilayah yang diperlukan oleh masyarakat dan lembaga-lembaga yang terkait dengan informasi tersebut. Variasi curah hujan yang tinggi belum ditunjang oleh sarana observasi yang memadai. Pada penelitian mencoba mengklasifikasi curah hujan kedalam tiga katagori yaitu : ringan, normal, dan deras. Pendekatan yang digunakan didasarkan pada metode pengklasifikasi *naive bayesian*. Penentuan klasifikasi didasarkan pada lima *feature* yaitu suhu minimum, suhu maksimum, kelembaban rata-rata, lama penyinaran, dan kecepatan angin. Data curah hujan yang digunakan sudah diketahui terlebih dahulu diamati kemudian diidentifikasi untuk membentuk dataset training. Selanjutnya dataset tersebut tersebut dijadikan dasar untuk melakukan penilaian untuk menentukan kelompok curah hujan. Berdasarkan hasil penelitian didapatkan tingkat akurasi klasifikasi curah hujan untuk katagori ringan,normal, dan deras berturut turut sebesar 79.5%, 40.9%, dan 86.4%.

**Kata Kunci : Pengklasifikasi, Naive bayesian, Curah hujan, Dataset**

### ***Abstract***

*Indonesia region is part of tropical region with high rainfall intensity. In some areas in Indonesia especially the area of Bali is often emerging a natural phenomenon that when the rainy season comes a lot of water abundance, based on annual data from BMKG (Meteorology Climatology and Geophysics Agency) rainfall data used to be done classification in order to provide information about the pattern of rainfall in a region that is needed by the community and the institutions associated with the*

*information. High variations in rainfall have not been supported by adequate observation tools. In the study tried to classify rainfall into three categories namely: mild, normal, and profusely. The approach used is based on the naive bayesian classifier method. Determination of classification is based on five features: minimum temperature, maximum temperature, average humidity, radiation length, and wind speed. The precipitation data used is known to be observed and then identified to form the training dataset. Furthermore, the dataset is used as the basis for conducting an assessment to determine the rainfall group. Based on the research results obtained accuracy level of rainfall classification for light, normal, and consecutive categories of 79.5%, 40.9%, and 86.4%.*

***Keywords : Classification, Naive Bayes, Rain Intensity, Dataset***

## **PENDAHULUAN**

Wilayah Indonesia merupakan bagian wilayah tropis dengan intensitas curah hujan yang tinggi. Hal ini dikarenakan adanya perbedaan lintang, gerak semua matahari, letak geografis, topologi serta interaksi berbagai macam sirkulasi udara baik itu lokal, regional maupun global. Variasi curah hujan yang tinggi belum ditunjang oleh sarana observasi yang memadai. Masih banyak lokasi terpencil yang miskin informasi cuaca dan iklimnya, padahal informasi ini penting. Pada beberapa wilayah di Indonesia khususnya wilayah Bali sering muncul suatu fenomena alam yaitu bila saat musim hujan tiba terjadi limpahan air yang cukup banyak, bahkan sampai menimbulkan bencana banjir. Namun sebaliknya, bila musim kemarau tiba ketersediaannya menjadi terbatas dan sering menimbulkan kekeringan bagi tumbuhan.

Cuaca adalah keadaan udara pada saat tertentu dan di wilayah tertentu yang relative sempit dan pada jangnan waktu yang singkat (Lakitan dan Benyamin,1997). Cuaca itu terbentuk dari gabungan unsur cuaca dan jangka waktu cuaca dalam waktu beberapa jam saja. Iklim adalah keadaan cuaca rata-rata dalam waktu satu tahun yang menyelidikannya dilakukan dalam waktu yang lama (minimal 30 tahun) dan meliputi wilayah yang luas (Lakitan dan Benyamin,1997). Ada beberapa unsur yang mempengaruhi cuaca dan iklim, yaitu suhu udara, tekanan udara, kelembaban udara dan curah hujan.

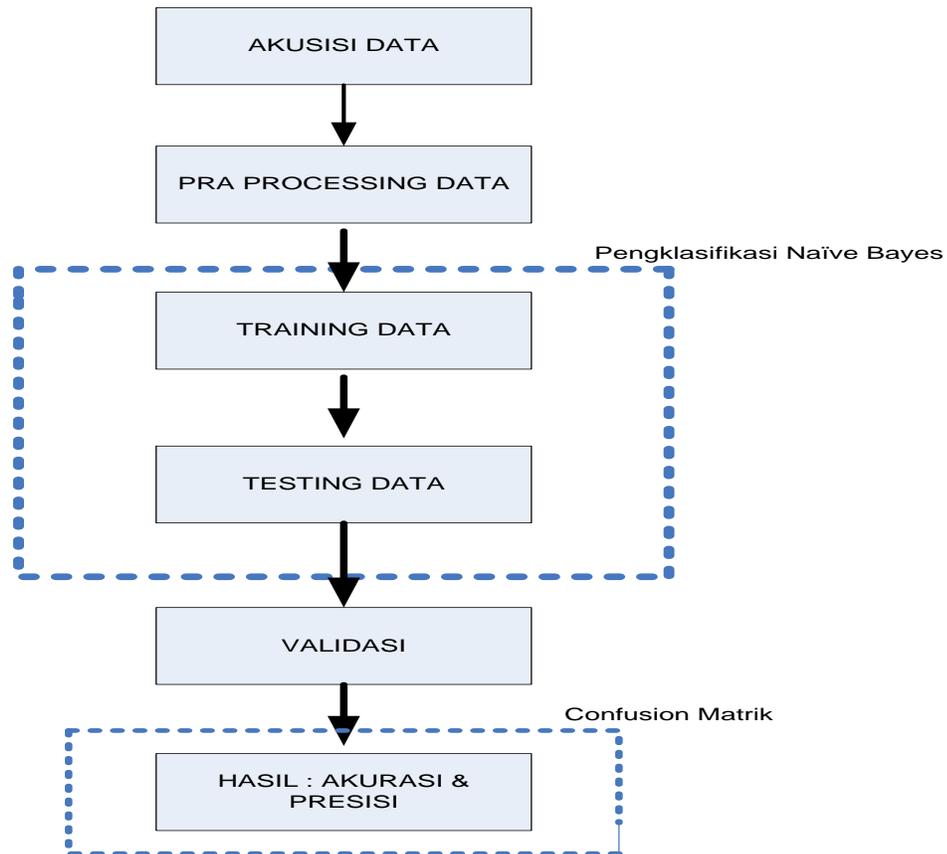
Penggunaan data satelit merupakan solusi yang banyak digunakan dalam rangka mengurangi kesenjangan informasi cuaca dan iklim tersebut. Data iklim yang digunakan merupakan data tahunan dari BMKG (Badan Meteorologi Klimatologi dan Geofisika). Namun untuk mendapatkan data satelit merupakan hal yang sulit, dan mahal. Pada sisi lain terdapat beberapa data yang bisa didapatkan pada stasiun pengamatan yang ada di bumi. Data tersebut seperti suhu, kecepatan angin, kelembaban, dan lain. Merupakan hal yang sangat berguna apabila dengan data iklim yang mudah didapat tersebut dapat digunakan untuk menganalisis keadaan cuaca.

Berdasarkan data BMKG, didapatkan beberapa data yang diyakini merupakan parameter atau *feature* untuk dapat menentukan keadaan cuaca (curah hujan). Beberapa *feature* yang dimaksud antara lain suhu minimum, suhu maksimum, kelembaban rata-rata, lama penyinaran, dan kecepatan angin.

Pada penelitian ini akan dicoba menentukan klasifikasi curah hujan berdasarkan lima *feature* diatas. Metode klasifikasi yang digunakan berdasarkan pendekatan *naive bayes*. Pengklasifikasi *naive bayes* merupakan pendekatan berdasarkan probabilitas, dalam beberapa penelitian merupakan pendekatan yang cukup baik digunakan sebagai metode klasifikasi.

## **METODE**

Metode penelitian untuk menentukan klasifikasi curah hujan ini dilakukan dalam beberapa tahapan, yaitu akusisi data, pengolahan data awal (*pra processing*) data, pelatihan (*training*) , pengujian (*testing*) dan validasi data, secara keseluruhan dinyatakan dalam bentuk diagram alur pada Gambar 1.



**Gambar 1. Alur Proses Penelitian**

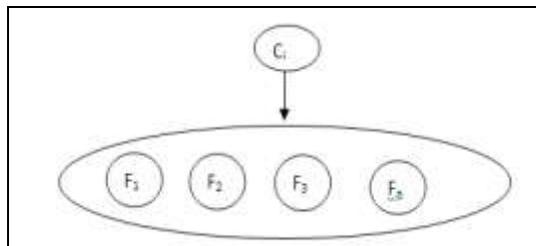
### **Akuisi Data**

Pembentukan data set untuk data training dilakukan dengan pemilihan data iklim tahunan dari 3 stasiun pengamatan BMKG di wilayah Bali dengan masing-masing data mendukung kategori curah hujan. Dipilih 150 data iklim dengan pembagian 50 data untuk kategori curah hujan ringan, 50 data untuk kategori curah hujan normal dan 50 data untuk kategori curah hujan deras, data tersebut dijadikan data training. Sedangkan untuk data testing digunakan 132 data, masing-masing terdiri atas 44 data , untuk hujan ringan, sedang, dan deras.

### Klasifikasi Curah Hujan Dengan Naive Bayesian

Klasifikasi dapat didefinisikan dengan detail sebagai suatu kegiatan melakukan pembelajaran atau pelatihan terhadap fungsi  $f$  yang memetakan vektor  $x$  ke dalam suatu dari beberapa label  $y$  kelas yang tersedia. Kegiatan tersebut akan memberikan hasil suatu model yang selanjutnya disimoan sebagai model (Prasetyo,2014). Model yang dibuat pada saat pembelajaran selanjutnya dapat digunakan untuk menentukan label kelasnya.

*Naïve Bayes Classifier* (NBC) juga disebut sebagai *Bayesian Classification* adalah metode pengklasifikasian statistic yang berguna untuk proses menentukan probabilitas suatu keanggotaan dari suatu class. Teorema *Bayes* mendasari *Naïve Bayes Classifier* yang mempunyai kemampuan klasifikasi yang serupa dengan *Decision Tree* dan *Neural Network*. NBC juga efisien, efektif dan handal menangani data seperti atribut yang tidak relevan. NBC juga bisa mengatasi *dataset* besar baik dengan atribut variable maupun kontinu (Yang, 2003) , (Luthfi,2019). NBC secara matematis dinyatakan dengan Persamaan 1. Proses perhitungan klasifikasi dengan metode *Naïve Bayesian* didasarkan pada perhitunagn peluang . Hasil klasifikasi dapat berupa kelas  $(C = C_1, C_2, \dots, C_i)$ , masing -masing kelas  $C_i$  ditentukan probabilitasnya didasarkan kemunculan beberapa *feature*  $F$ , dimana  $F = F_1, F_2, \dots, F_n$ . Uraian ini diilustrasikan pada Gambar 2.



Gambar 2. Ilustrasi klasifikasi Bayes

Perhitungan klasifikasi setiap kelas ( $C_i$ ) berdasarkan metode *naive bayes* dinyatakan pada Persamaan 1, dan selanjutnya dengan penurunan lebih lanjut didapatkan Persamaan 2.

$$P(C_i|F) = \frac{P(F|C_i) \cdot P(C_i)}{P(x)} \dots\dots\dots(1)$$

$$p(C|F_1, F_2, \dots, F_n) = p(c) \cdot (p(F_1|C) \cdot \dots \cdot p(F_n|C)) \dots\dots\dots(2)$$

Pada data dengan nilainya berupa data diskrit maka penentuan  $P(F|C)$  ditentukan dengan menghitung peluang yang didasarkan pada frekuensi relatif biasa. Sebaliknya jika datanya berupa data kontinyu, maka  $P(F|C)$  ditentukan berdasarkan distribusi normal (*gaussian*), dengan cara menghitung nilai dari rerata  $\mu$  serta standar deviasi  $\sigma$  di setiap *class* dinyatakan pada Persamaan 3, (Prasetyo, 2014) :

$$P(F|C) = \frac{1}{\delta \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\delta^2}} \dots\dots\dots(3)$$

Berurut  $\mu$  dan  $\delta$  adalah nilai mean dan deviasi standar dari *feature* F pada kelas C.

**Validasi Pengukuran**

Validasi dilakukan dengan membandingkan kelas data pada kondisi asal (originalnya), dengan hasil kelas yang diputuskan oleh sistem. Terdapat empat kemungkinan, yaitu data asal dalam kondisi benar dan hasil prediksi sistem benar, kondisi disebut *True positive*(TP). Kondisi kedua disebut dengan *false negative* (FN) yaitu data asal dalam katagori benar, namun hasil prediksi salah. Kondisi ketiga disebut *False positive* (FP), data salah namun diprediksi benar, dan kondisi keempat yaitu data salah dan diprediksi salah disebut dengan *true negative* (TN) (Wahyuni,dkk, 2014).

*Confusion matrix* yaitu mempresentasikan hasil validasi sistem dengan menggunakan tabel matriks, dinyatakan pada Gambar 3.

		Kelas hasil prediksi (j)	
		Kelas = 1	Kelas = 0
Kelas asli (i)	Kelas = 1	TP	FN
	Kelas = 0	FP	TN

**Gambar 3. Confusion Matrik**

Berdasarkan matrik konfusi dapat ditentukan beberapa parameter untuk menilai kinerja sistem, diantaranya akurasi, presisi, *recall*, laju kesalahan [7]. Tujuan dari pengukuran akurasi adalah untuk mengetahui kinerja dari *Naïve Bayesian* dalam menentukan klasifikasi cuaca ke dalam kelas yang telah ditentukan. Penentuan akurasi, presisi, *recall*, laju kesalahan dinyatakan pada Persamaan 4,5,6,7.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(4)$$

$$Presisi = \frac{TP}{TP+FP} \dots\dots\dots(5)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(6)$$

$$Laju\ Error = \frac{FN+FP}{TP+TN+FN+FP} \dots\dots\dots(7)$$

Nilai akurasi adalah persentase jumlah record data yang diklasifikasikan dengan benar oleh suatu algoritma klasifikasi ( Han dkk, 2012) . Nilai *recall* digunakan untuk membandingkan *True Positive* dengan tupel positif. Sedangkan nilai presisi adalah prosisi kasus dengan hasil diagnose positif (Wahyuni dkk, 2014).

## HASIL DAN PEMBAHASAN

Penelitian dilakukan dalam beberapa langkah mulai dari pembentukan dataset, perhitungan parameter statistik dan pengujian.

### Pembentukan Dataset

Dilakukan penentuan kriteria F<sub>1</sub>:suhu minimum, F<sub>2</sub>:suhu maksimum, F<sub>3</sub>:kelembaban rata-rata, F<sub>4</sub>:lama penyinaran, dan F<sub>5</sub>:kecepatan angin rata-rata. Dalam melakukan penelitian ini diperlukan data iklim sebanyak 150 data yakni data iklim dari bulan Januari 2012 sampai dengan Desember 2017 yang digunakan sebagai dataset. Dataset tersebut didapatkan dari situs <http://dataonline.bmkg.go.id>. Berdasarkan dataset yang didapatkan, selanjutnya dibagi menjadi 3 bagian sesuai dengan kategori curah hujan yang akan diklasifikasi dengan komposisi 50 dataset untuk masing-masing kategori.

Pada dataset yang didapat kriteria F<sub>1</sub> sampai F<sub>6</sub> adalah data kontinu sehingga probabilitas dihitung berdasarkan desintas gauss. Hasil perhitungan probabilitas untuk semua kriteria ditunjukkan pada tabel 1.

**TABEL 1. DATA TRAINING**

Kategori Non Verbal	Nilai rata-rata ( $\mu$ ) dan standard deviasi ( $\delta$ )		
	Ringan	Normal	Deras
F <sub>1</sub> :Suhu minimum	$\mu = 24.94$	$\mu = 23.88$	$\mu = 23.69$
	$\delta = 1.35$	$\delta = 0.75$	$\delta = 0.815$
F <sub>2</sub> :Suhu maksimum	$\mu = 31.41$	$\mu = 30.46$	$\mu = 30.15$
	$\delta = 1.289$	$\delta = 1.09$	$\delta = 1.98$
F <sub>3</sub> :Kelembaban rata-rata	$\mu = 80.38$	$\mu = 84.84$	$\mu = 86.2$
	$\delta = 4.42$	$\delta = 4.15$	$\delta = 4.89$
F <sub>4</sub> :Lama penyinaran	$\mu = 5.99$	$\mu = 4.21$	$\mu = 2.39$
	$\delta = 3.75$	$\delta = 3.34$	$\delta = 2.74$
F <sub>5</sub> :Kecepatan angin rata-rata	$\mu = 6.92$	$\mu = 7.68$	$\mu = 7.72$
	$\delta = 2.99$	$\delta = 3.99$	$\delta = 3.92$

### Hasil Pengujian dan Pembahasan

Hasil perhitungan statistik peluang pada tabel 1 selanjutnya digunakan untuk menghitung derajat atau probabilitas kategori curah hujan. Sebagai testing digunakan 132 data iklim yang sudah diketahui validitas kategori curah hujannya. Tabel 2 dan 3 menunjukkan data *testing* hasil klasifikasi yang mewakili masing-masing kategori hujan. , sedangkan tabel 4 menunjukkan hasil perhitungan klasifikasi curah hujan ringan, normal dan deras dengan Naïve Bayesian.

**TABEL 2. HASIL PERHITUNGAN DATA ASLI KATAGORI BENAR**

		HASIL KLASIFIKASI		
		RINGAN	NORMAL	DERAS
DATA ASLI	RINGAN	27	15	2
	NORMAL	16	23	5
	DERAS	0	4	40

**TABEL 3. HASIL PERHITUNGAN DATA ASLI KATAGORI SALAH**

		HASIL KLASIFIKASI		
		RINGAN	NORMAL	DERAS
DATA ASLI	RINGAN	43	1	0
	NORMAL	1	13	30
	DERAS	0	8	36

Secara keseluruhan kinerja klasifikasi didapatkan nilai akurasi curah hujan diprovinsi Bali, dinyatakan pada Tabel 4

**TABEL 4. HASIL KINERJA KLASIFIKASI CURAH HUJAN**

PARAMETER	KATAGORI		
	RINGAN	NORMAL	DERAS
AKURASI	79.5%	40.9%	86.4%
PRESISI	96.4%	42.6%	83.3%
<i>RECALL</i>	61.4%	52.3%	90.9%
<i>ERROR RATE</i>	20.5%	59.1%	13.6%

Dari pengujian 132 data, masing-masing katagori hujan, ringan, normal dan deras 88 data. Dari hasil pengujian didapatkan nilai akurasi untuk klasifikasi hujan dengan katagori ringan, normal, dan deras berturut turut 79.5%, 40.5% dan 86.4%. Berdasarkan hal tersebut klasifikasi pada kondisi hujan normal, menunjukkan nilai akurasi yang rendah. Hal ini menunjukkan terjadi kesulitan dalam menentukan deteksi cuaca normal atau bukan normal. Berdasarkan hasil analisis pada data *training*, tidak ditemukan ada perbedaan yang signifikan diantara ketiga katagori hujan. Dengan demikian tidak dapat dinyatakan pendekatan *naive bayes* untuk klasifikasi curah hujan gagal. Dengan demikian hal yang perlu dicoba adalah melakukan pengujian ulang dengan menambah data testing, terutama pada data curah hujan normal.

## **KESIMPULAN DAN SARAN**

### **Kesimpulan**

Dari hasil pengujian yang telah dilakukan dapat disimpulkan beberapa hal sebagai berikut:

1. Dari hasil pengujian menunjukkan bahwa klasifikasi kategori curah hujan dengan katagori ringan, normal, dan deras berturut - turut adalah 79.5%, 40.9 %, dan 86.4 % . Sedangkan kepresisiannya berturut -turut 96.4%, 42.6%, dan 83.3%.

2. Tingkat akurasi yang rendah 40.9% dan presisi 42.6% pada klasifikasi curah hujan pada kelas Normal, menunjukkan sistem belum bisa menghasilkan hasil klasifikasi yang baik. Terjadi ketidak stabilan dalam pengukuran, dinyatakan dengan nilai akurasi yang rendah tersebut.
3. Klasifikasi pada kategori curah hujan ringan, dan deras, sistem sudah dapat mengklasifikasi dengan baik.
4. Kemungkinan hasil yang masih rendah pada kategori hujan normal terletak pada data, seharusnya tidak ada perbedaan yang signifikan diantara ketiga kelas. Perlu dilakukan pengujian ulang dengan data yang lebih banyak.

### **Saran**

Berdasarkan hasil penelitian pembahasan dan analisa pada penelitian, maka peneliti memberikan saran untuk penelitian selanjutnya yang diharapkan bias dijadikan dasar untuk mengembangkan penelitian ini yaitu:

1. Perlu dilakukan pengujian ulang dengan metode yang sama dengan jumlah data yang lebih banyak, pada penelitian ini hasil klasifikasi pada kategori curah hujan normal masih rendah. Salah satu kemungkinan masalah adalah pada kualitas dan kuantitas data.
2. Perlu adanya penelitian pada tahap selanjutnya, yaitu dengan menggunakan metode lain atau menggabungkan beberapa metode sehingga diharapkan dapat meningkatkan akurasi dengan menutupi kekurangan metode yang telah digunakan sebelumnya.

### **DAFTAR PUSTAKA**

- D. T. Wahyuni, T. Sutojo dan A. Luthfiarta, "Prediksi DKI Jakarta Menggunakan Naïve Bayes dengan Algoritma Genetika Sebagai Fitur Seleksi", 2014.
- E. Prasetyo, DATA MINING – Mengolah Data Menjadi Informasi Menggunakan Matlab, Yogyakarta: ANDI, 2014.

I. H. Written, E. Frank dan M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques*, Burlington: Morgan Kaufmann, 2011.

Lakitan. Benyamin, 1997, *Dasar Dasar Klimatologi*. Raja Grafindo Persada, Jakarta

Y Yang and G. I. Webb, "On Why Discretization Works for Naïve-Bayes Classifier", 2003.

K. dan E. T. Luthfi, *Algoritma Data Mining*, Yogyakarta: Penerbit Andi, 2009.

J. Han, M. Kamber dan J. Pei, *Data Mining Concepts and Techniques*, Waltham: Morgan Kaufmann, 2012.

<http://dataonline.bmkg.go.id/>