

## Analysis of the Naïve Bayes Classifier Method in Classifying the Weather Conditions in Aceh Tamiang

Defy Syafira<sup>1\*</sup>

<sup>1</sup> Universitas Samudra, Indonesia

\*Corresponding Author Email: [\\*defysyafira@gmail.com](mailto:defysyafira@gmail.com)

### ABSTRAK

**Received: 30 March 2024**  
**Revised: 31 March 2024**  
**Accepted: 31 March 2024**  
**Available online: 1 April 2024**

#### **Kata Kunci:**

*Naive Bayes Classifier, Cuaca, Prediksi*

Naive Bayes Classifier adalah salah satu metode yang bertujuan untuk memprediksi/meramalkan kondisi masa depan. Metode ini merupakan salah satu sub bagian dari pengklasifikasian. Pada penelitian ini, pengujian yang dilakukan untuk memprediksi kondisi cuaca pada masa tertentu. Dari hasil pengujian yang dilakukan terhadap 4 kondisi cuaca pada 3 kota yang berbeda dengan menggunakan metode Naive Bayes Classifier diperoleh hasil akurasi prediksi akurat sebesar 92,1%, sedangkan dengan menggunakan metode ARIMA (Autoregressive Integrated Moving Average) hasil akurasi yang diperoleh sebesar 86,8%. Hasil ini menunjukkan bahwa Naive Bayes Classifier memiliki tingkat persentase akurasi yang lebih besar dibandingkan ARIMA, yaitu sebesar 5.3%.

### ABSTRACT

#### **Keywords:**

*Naive Bayes Classifier, Weather, Prediction*

*Naive Bayes Classifier is a method that aims to predict future conditions. This method is a sub-part of classification algorithms. In this study, tests were tested out to predict weather conditions at a certain time. The results of the studies carried out on 4 weather conditions in 3 different cities in Indonesia using Naive Bayes Classifier method the accurate prediction accuracy is 92.1%, while using ARIMA (Autoregressive Integrated Moving Average) method the accuracy results obtained are 86.8 %. These results indicate that Naive Bayes Classifier has a greater percentage level of accuracy than ARIMA, which is 5.3%.*

## 1. INTRODUCTION

Algoritma klasifikasi kebanyakan menjadi bermasalah pada data dengan memiliki atribut yang banyak, contohnya dalam penurunan akurasi klasifikasi (Tan et al, 2006). Salah satu metode yang memungkinkan algoritma klasifikasi untuk bekerja lebih cepat dan efektif serta meningkatkan akurasi dan performansi suatu algoritma klasifikasi adalah dengan mereduksi atributnya (Prasetyo, E, 2014). Reduksi atribut dapat menghilangkan fitur yang tidak relevan, mengurangi noise, dan mengurangi curse of dimensionality. Reduksi atribut juga dapat mengurangi jumlah waktu dan memori yang dibutuhkan oleh algoritma klasifikasi.

Naïve Bayes adalah sebuah metode klasifikasi yang berdasarkan probabilitas dan teorema Bayesian dengan asumsi keindependenan atribut. Asumsi independensi atribut akan menghilangkan kebutuhan banyaknya jumlah data latih dari seluruh atribut yang dibutuhkan untuk mengklasifikasi suatu data. Data latih untuk Teorema Bayes membutuhkan paling tidak perkalian kartesius dari seluruh kelompok atribut yang mungkin, sehingga semakin sedikit atribut yang digunakan, akan mengurangi data latih yang dibutuhkan. Namun, pada kenyataannya asumsi atribut independen pada Naïve Bayes Classifier sering diabaikan. Hal ini disebabkan karena asumsi

keindependenan atribut dalam dunia nyata hamper tidak pernah terjadi.

Babajide (2017), pada penelitiannya menggunakan Naïve Bayes Classifier untuk memprediksi tingkat hipertensi. Tingkat akurasi yang didapatkan dalam penelitian tersebut adalah sebesar 83,67% untuk Naïve Bayes sedangkan dengan menggunakan Decision Trees diperoleh tingkat akurasi sebesar 77,55%.

Handayani (2015) didalam penelitian membahas implementasi algoritma Naïve Bayes Classifier untuk pengklasifikasian teks otomatis pengaduan dan pelaporan masyarakat melalui layanan callcenter 110. Berdasarkan hasil penelitian tersebut, diperoleh tingkat akurasi untuk recall sebesar 93%, precision sebesar 93%, f-measure sebesar 92% dan rata-rata akurasi sebesar 92,67%.

Yofi (2018), dalam penelitiannya menggunakan metode Naïve Bayes Classifier dan K-Nearest Neighbor untuk mengklasifikasi Kartu Indonesia Sehat (KIS). Hasil yang diperoleh pada penelitian tersebut adalah tingkat akurasi sebesar 96% dengan mengkombinasikan metode Naïve Bayes pada algoritma K-Nearest Neighbor.

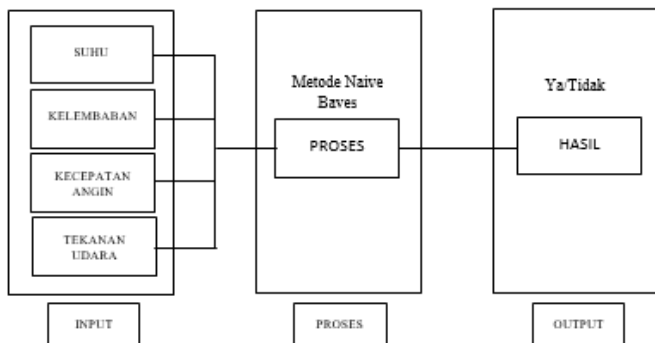
Sementara pada penelitian ini, pengklasifikasian kondisi cuaca akan menggunakan metode Naïve Bayes Classifier terhadap metode ARIMA (Autoregressive Integrated Moving Average). Untuk membuktikan bahwa ada perbedaan dalam

sistem penyelesaian metode tersebut, maka akan diukur tingkat akurasinya untuk proses pengklasifikasiannya.

## 2. RESEARCH METHODS

Data yang digunakan pada penelitian ini adalah beberapa data yang diperoleh dari berbagai sumber. Adapun data cuaca yang diambil adalah data yang disediakan oleh sebuah penyedia informasi cuaca yaitu meteoblue.com. Data yang didapat dari cloud server meteoblue.com tersebut kemudian diolah untuk dikelaskan sesuai dengan kriteria yang telah ditentukan. Contohnya suhu, pada penelitian ini dibagi menjadi tiga jenis kelas, yaitu suhu rendah, suhu sedang dan suhu tinggi. Begitu juga dengan beberapa atribut lainnya, masing-masing dikelaskan sesuai dengan spesifikasinya. Setelah proses pembentukan kelas variabel ini, kemudian dilakukan pengolahan data dengan menggunakan metode Naïve Bayes Classifier untuk menentukan data cuaca yang didapat dari cloud untuk wilayah yang ditentukan termasuk kedalam kondisi nyaman maupun tidak nyaman.

### 2.1 Arsitektur Algoritma yang Diusulkan

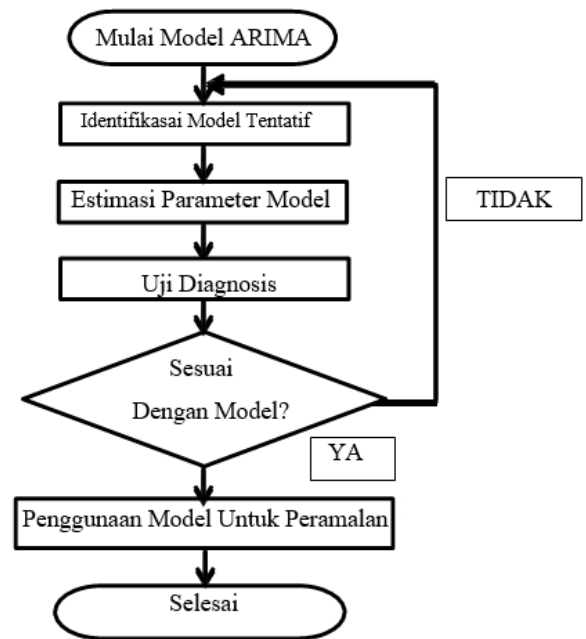


Gambar 1. Langkah Penelitian

Berdasarkan gambar 1, langkah pertama yang akan dilakukan adalah menginput dataset sebagai bahan yang akan digunakan dalam pengolahan data. Setelah itu maka dataset tersebut kemudian akan digunakan untuk pengklasifikasian menggunakan algoritma Naive Bayes Classifier. Adapun 4 atribut yang diinput adalah suhu, kelembaban, kecepatan angin, dan tekanan udara. Setelah di input, selanjutnya melakukan tahapan proses pengumpulan data suhu, kelembaban udara, kecepatan angin dan tekanan udara yang berguna sebagai data training, di lanjutkan pengkategorian data skala suhu dingin /sedang/ panas, kelembaban nyaman/lembab, kecepatan angin cepat/standar dan tekanan udara normal/tidak normal, dan selanjutnya yaitu perhitungan peluang kejadian masing-masing cuaca.

Setelah dilakukan pengklasifikasian maka akan dihitung tingkat akurasinya dan menganalisisnya sehingga nanti akan ditarik kesimpulan dari hasil penelitian yang nanti akan dilakukan.

## 2.2 Skema Algoritma



Gambar 2. Skema Algoritma

Berdasarkan gambar 2, langkah pertama yang akan dilakukan adalah mengidentifikasi model tentatif yang kemudian dilanjutkan dengan mengestimasi parameter modelnya. Kemudian setelah mengestimasi maka dilakukan pengujian diagnosis model, jika model telah sesuai maka langkah selanjutnya dilakukan perhitungan model tersebut untuk pengklasifikasian menggunakan metode ARIMA.

Setelah dilakukan pengklasifikasian maka akan dihitung tingkat akurasinya dan menganalisisnya sehingga nanti akan ditarik kesimpulan dari hasil penelitian yang nanti akan dilakukan.

## 3. RESULT AND DISCUSSION

Pada penelitian ini, dengan menggunakan metode Naive Bayes Classifier untuk menentukan kelas untuk tiap data, dimana data yang dimaksud adalah data cuaca untuk kota atau wilayah yang di input dari meteoblue.com. Data tersebut kemudian akan dikalkulasikan dengan menggunakan data training Naive Bayes Classifier.

Dalam penelitian ini digunakan empat buah variabel data cuaca yaitu:

1. Suhu.

Dikarenakan data suhu yang didapat dari meteoblue.com adalah data asli dari nilai pengukuran suhu udara dalam bentuk derajat celcius, maka penulis melakukan konversi data tersebut menjadi kelas masing-masing dengan rincian sebagai berikut:

- Suhu Dingin, antara 0° sampai dengan 20°.
- Suhu Sedang, antara 21° sampai dengan 30°.
- Suhu Tinggi, untuk suhu diatas 31°

2. Kelembaban Udara

Data kelembaban udara yang diperoleh dari meteoblue.com juga merupakan data asli dalam bentuk persentase kelembaban udara dengan rumus:

$$RH = \frac{P_w}{P_{ws}(t)} \times 100\%$$

Maka dalam penelitian ini, penulis juga melakukan konversi nilai asli tersebut menjadi kelas masing-masing dengan rincian sebagai berikut:

- Kelembaban Kering, dimana pada kondisi ini nilai dari *Relative Humidity* adalah mulai dari 0% sampai dengan 44%.
- Kelembaban Nyaman, dimana pada kondisi ini nilai dari *Relative Humidity* adalah berkisar antara 45% sampai dengan 75%.
- Kelembaban Tinggi, dimana *Relative Humidity* pada kondisi ini adalah berkisar diatas 75%.

3. Kecepatan Angin.

Data kecepatan angin yang diperoleh dari meteoblue.com juga merupakan data asli berupa nilai kecepatan udara yang dihitung dalam satuan meter per detik (m/s). Oleh karena itu, penulis juga melakukan konversi data Kecepatan Angin menjadi kelas masing-masing dengan rincian sebagai berikut:

- Kecepatan Angin Normal, dimana pada kondisi ini nilai daripada kecepatan angin berkisar dari 0.5 m/s sampai dengan 4 m/s.
- Kecepatan Angin Cepat, dimana kondisi ini digunakan untuk kecepatan angin diatas 4 m/s.

4. Tekanan Udara.

Data tekanan udara yang diperoleh dari meteoblue.com juga berupa nilai asli dari nilai yang didapatkan oleh barometer (alat pengukur tekanan udara) yang dinyatakan dalam hPa (*hektopascal*). Oleh karena itu, penulis juga melakukan konversi daripada nilai asli yang didapat dari cloud server meteoblue.com ini menjadi kelas masing-masing dengan rincian sebagai berikut:

- Tekanan Udara Normal, dimana kelas ini digunakan untuk kondisi tekanan udara yang bernilai 1006 hPa sampai dengan 1018 hPa.
- Tekanan Udara Tidak Normal, dimana kelas ini digunakan untuk kondisi tekanan udara yang bernilai dibawah 1006 hPa dan di atas 1018 hPa.

Dalam penelitian ini dilakukan pengujian data terhadap 3 wilayah yang ada di Provinsi Aceh, yaitu:

- Kota Kuala Simpang, Aceh Tamiang yang memiliki cuaca yang sangat bervariasi.
- Kota Medan, merupakan ibu kota Provinsi Sumatera Utara.
- Kota Jakarta, yang merupakan ibu kota Negara Indonesia.

Adapun data cuaca yang di ambil dari meteoblue.com dalam penelitian ini menjadi 4 bagian waktu seperti berikut ini:

- Dini hari, dengan rentang waktu dari jam 00.00 sampai dengan jam 06.00
- Pagi hari, dengan rentang waktu dari jam 06.00 sampai dengan jam 12.00
- Siang hari, dengan rentang waktu dari jam 12.00 sampai dengan jam 18.00
- Malam hari, dengan rentang waktu dari jam 18.00 sampai dengan jam 00.00

Adapun data yang akan diambil untuk digunakan adalah data selama 8 (delapan) hari berurutan, dimulai dari tanggal 1 April 2020 sampai tanggal 8 April 2020. Data tersebut kemudian diolah untuk mendapatkan hasil prediksi data cuaca untuk hari esoknya, yaitu tanggal 9 April 2020. Kemudian langkah terakhir yaitu menghitung data prediksi cuaca yang telah diolah untuk mengetahui tingkat kenyamanannya.

Adapun sampel data yang diperoleh dari situs meteoblue adalah sebagai berikut:

**Tabel 1.** Sampel Data

Tanggal	Waktu	temp	RH	press	wind
1	Dini	25,82	94,50	1013,10	0,90
	Pagi	27,21	83,00	1013,73	0,70
	Siang	32,71	63,00	1011,58	1,82
	Malam	24,45	82,67	1013,33	1,02
2	Dini	25,78	96,50	1012,67	0,95
	Pagi	28,10	82,67	1013,60	0,65
	Siang	31,78	59,33	1010,82	2,05
	Malam	28,24	76,50	1012,77	1,17
3	Dini	25,74	88,17	1011,67	1,03
	Pagi	27,12	80,33	1012,37	1,42
	Siang	29,25	73,50	1010,80	1,58
	Malam	27,97	83,17	1012,05	1,12
4	Dini	26,16	91,50	1011,03	1,23
	Pagi	28,42	78,33	1011,65	2,20
	Siang	29,79	74,00	1009,72	2,10
	Malam	27,54	84,17	1011,48	1,20
5	Dini	26,35	89,50	1010,33	1,12
	Pagi	29,16	76,00	1010,63	1,35
	Siang	30,94	69,17	1008,58	2,52
	Malam	27,78	85,50	1010,18	0,83
6	Dini	26,60	92,17	1010,12	0,98
	Pagi	29,21	73,33	1011,15	1,28
	Siang	32,79	58,50	1009,07	3,30
	Malam	28,26	84,17	1011,13	0,85
7	Dini	26,86	94,00	1010,95	1,03
	Pagi	29,24	76,50	1011,80	1,23
	Siang	30,06	73,67	1010,37	1,98
	Malam	27,75	88,83	1012,37	0,72
8	Dini	27,08	95,33	1011,75	0,68
	Pagi	28,62	81,50	1012,32	0,65
	Siang	29,05	76,33	1011,00	1,52
	Malam	27,31	86,50	1012,33	1,08

**3.1 Perhitungan Prediksi Kondisi Cuaca**

Dalam penelitian ini, dilakukan prediksi cuaca untuk tanggal 9 April 2020 dengan menggunakan data yang telah diperoleh sebelumnya dari meteoblue.com yang dimulai dari tanggal 1 sampai dengan tanggal 8 April 2020 dengan menggunakan metode kalkulasi rata-rata. Adapun hasil dari prediksi berdasarkan data asli dari meteoblue.com adalah seperti pada tabel dibawah ini:

**Tabel 2.** Prediksi Kondisi Cuaca

Waktu	Kondisi Cuaca			
	Temp	RH	Wind	Pressure

Dini Hari	26,17	92,69	1,08	1011,20
Pagi Hari	30,14	69,37	1,47	1012,18
Siang Hari	30,17	71,06	2,02	1009,72
Malam Hari	27,25	87,37	0,88	1012,55

### 3.2 Perhitungan Klasifikasi Data dengan Naïve Bayes

Adapun langkah-langkah perhitungan manual untuk proses klasifikasi data dengan menggunakan metode Naive Bayes Classifier adalah sebagai berikut:

1. Hitung probabilitas tiap-tiap kelas.

$$P(\text{Class}|\text{Baik}) = \frac{16}{38} = 0.421$$

$$P(\text{Class}|\text{Buruk}) = \frac{22}{38} = 0.578$$

2. Membuat *Frequency Table* dari data training.

Langkah pertama yang dilakukan yaitu menghitung probabilitas kemunculan untuk tiap-tiap variabel cuaca berdasarkan kelas

**Tabel 3.** *Frequency Table* Suhu

Frequency Table		Kondisi	
		Baik	Buruk
Suhu	Dingin	5	8
	Sedang	5	8
	Panas	6	6

**Tabel 4.** *Frequency Table* Kelembaban Udara

Frequency Table		Kondisi	
		Baik	Buruk
Kelembaban Udara	Kering	3	9
	Nyaman	9	5
	Lembab	4	8

**Tabel 5.** *Frequency Table* Kecepatan Angin

Frequency Table		Kondisi	
		Baik	Buruk
Kecepatan Angin	Normal	11	8
	Cepat	5	14

**Tabel 6.** *Frequency Table* Tekanan Udara

Frequency Table		Kondisi	
		Baik	Buruk
Tekanan Udara	Normal	13	6
	Tidak	3	16

3. Menghitung probabilitas dan tingkat kenyamanan cuaca di tiap kota untuk 4 waktu (dini hari, siang hari, sore hari dan malam hari).

- a) Perkiraan Cuaca di Kota Kuala Simpang Tanggal 9 April 2020 adalah:

Dini Hari

kondisi "Baik"

$$P(X|C)$$

$$P(\text{kondisi}|Y) = \frac{16}{38}$$

$$P(\text{Sedang}|Y) = \frac{5}{16}$$

$$P(\text{Lembab}|Y) = \frac{4}{16}$$

$$P(\text{Standar}|Y) = \frac{11}{16}$$

$$P(\text{Normal}|Y) = \frac{13}{16}$$

$$(P|Y) = \frac{16}{38} \times \frac{5}{16} \times \frac{4}{16} \times \frac{11}{16} \times \frac{13}{16} = 0,0183747944$$

kondisi "Buruk"

$P(X|C)$

$$P(\text{kondisi}|N) = \frac{22}{38}$$

$$P(\text{Sedang}|N) = \frac{8}{22}$$

$$P(\text{Lembab}|N) = \frac{8}{22}$$

$$P(\text{Standar}|N) = \frac{6}{22}$$

$$P(\text{Normal}|N) = \frac{6}{22}$$

$$(P|Y) = \frac{22}{38} \times \frac{8}{22} \times \frac{8}{22} \times \frac{8}{22} \times \frac{6}{22} = 0,0075922338$$

### 3.3 Pengujian Sistem Dengan Confusion Matrix

Confusion matrix merupakan suatu metode yang digunakan untuk menghitung akurasi dan menguji kinerja dari sistem klasifikasi (Hay, 1988). Rumus perhitungan akurasi yaitu

$$\text{Accuracy} = \frac{TP + TN}{n}$$

Adapun perhitungan confusion matrix untuk sistem klasifikasi yang dibangun ini adalah sebagai berikut:

**Tabel 7.** *Confusion Matrix*

Confusion Matrix		Prediksi	
		Buruk	Baik
Aktual	Buruk	26	0
	Baik	3	9

Adapun perhitungan manual untuk confusion matrix adalah sebagai berikut:

$$\text{Accuracy} = \frac{26 + 9}{38} = 0,921 = 92,1 \%$$

## 4. CONCLUSION

Berdasarkan pengujian yang telah dilakukan maka penulis dapat mengambil beberapa kesimpulan yang diuraikan sebagai berikut:

1. Tingkat Akurasi sistem Prediksi cuaca menggunakan *Naive Bayes Classifier* adalah sebesar 92.1%.
2. Dari 12 prediksi yang menyatakan kondisi baik, 9 diantaranya memiliki hasil yang sama dengan data *real*. Dan dari 26 prediksi yang menyatakan kondisi buruk, keseluruhannya memiliki hasil yang sama dengan data *real*. Hal ini menunjukkan bahwa prediksi cuaca menggunakan *Naive Bayes Classifier* dapat

dikembangkan lagi untuk peningkatan akurasi prediksinya menggunakan bantuan metode lainnya.

## REFERENCES

- [1] Arifin, M. 2015. Ig-Knn Untuk Prediksi Customer Churn Telekomunikasi. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 6(1), 1-10
- [2] Babajide, O. A., Thomas, I.A., & Iyanuoluwa, A.O. (2017). Hypertension Prediction System Using Naïve Bayes Classifier. *JAMCS*. Vol 24, No 2. 1-11.
- [3] Dinata, R. K., Bustami., Fiasari., & Retno, S. (2023). Z-Score and Floyd Warshall Algorithms for Determining Alternative Routes of Mugging-Prone Areas in Medan City, Indonesia. *ILKOM Jurnal Ilmiah*, 15(3), 436-444.
- [4] Handayani, F. & Pribadi, S.F. (2015). Implementasi Algoritma Naïve Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*. Vol 7 No 1. 19-24.
- [5] Novakovic, Jasmina. 2010. The Impact of Feature Selection on the Accuracy of Naive Bayes Classifier. Vol 2. 113–116.
- [6] Prasetyo, E. 2014. Reduksi Dimensi Set Data dengan DRC pada Metode Klasifikasi SVM dengan Upaya Penambahan Komponen Ketiga. *Prosiding SNATIF*, 293-300.
- [7] Silwattananusarn, T & Tuamsuk, K. 2012. Data Mining and Its Applications for Knowledge Management : A Literature Review from 2007 to 2012. *IJDKP*. Vol 2, No 5. 13-24.
- [8] Taheri, S. & Mammadov, M. (2013). Learning The Naïve Bayes Classifier With Optimization Models. *IJAMCS*. Vol 23, no 4. 787-795.
- [9] Yofi, F. S., Riza, A., & Much, A.M. (2018). K-Nearest Neighbor and Naïve Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor. *SJI*. Vol 5 No 1. 9-18.