



**PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFIER DALAM  
KLASIFIKASI CURAH HUJAN DKI JAKARTA**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2022**



**PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFIER DALAM  
KLASIFIKASI CURAH HUJAN DKI JAKARTA**



*Tugas Akhir*

UNIVERSITAS  
MERCU BUANA  
Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:  
AHMAD SYAUQI IRHAMNA

41518010034

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2022

## LEMBAR PERNYATAAN ORISINALITAS

### LEMBAR PERNYATAAN ORISINALITAS

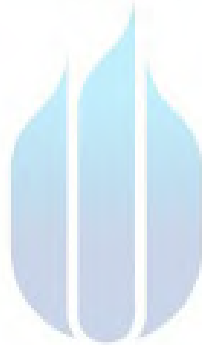
Yang bertanda tangan dibawah ini:

NIM : 41518010034

Nama : Ahmad Syauqi Irhamna

Judul Tugas Akhir : Penerapan Algoritma Naïve Bayes Classifier Dalam  
Klasifikasi Curah Hujan DKI Jakarta

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



Jakarta, 07 Juli 2022



Ahmad Syauqi Irhamna

UNIVERSITAS  
MERCU BUANA

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

### SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Ahmad Syauqi Irahmna  
NIM : 41518010034  
Judul Tugas Akhir : Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Curah Hujan DKI Jakarta

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 07 Juli 2022

  
Ahmad Syauqi Irahmna

## SURAT PERNYATAAN LUARAN TUGAS AKHIR

### SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Ahmad Syauqi Irhamna  
NIM : 41518010034  
Judul Tugas Akhir : Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Curah Hujan DKI Jakarta

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan ✓
		Jurnal Nasional Terakreditasi ✓	
		Jurnal Internasional Tidak Bereputasi	Diterima
		Jurnal Internasional Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Informatika: Jurnal Pengembangan IT	
	ISSN	: 2477-5126 (print), 2548-9356 (online)	
	Link Jurnal	: <a href="https://ejournal.poltektegal.ac.id/index.php/informatika">https://ejournal.poltektegal.ac.id/index.php/informatika</a>	
	Link File Jurnal Jika Sudah di Publish		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 07 Juli 2022



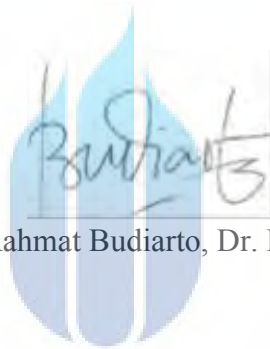
Ahmad Syauqi Irhamna

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010034  
Nama : Ahmad Syauqi Irhamna  
Judul Tugas Akhir : Penerapan Algoritma Naïve Bayes Classifier  
Dalam Klasifikasi Curah Hujan DKI Jakarta

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 27 Juli 2022



(Rahmat Budiarto, Dr. Prof)

UNIVERSITAS  
MERCU BUANA

## LEMBAR PERSETUJUAN PENGUJI

### LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010034  
Nama : Ahmad Syauqi Irhamna  
Judul Tugas Akhir : Penerapan Algoritma Naïve Bayes Classifier  
Dalam Klasifikasi Curah Hujan DKI Jakarta

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 27 Juli 2022



(Rushenda, S.Kom, MT)

UNIVERSITAS  
MERCU BUANA



## LEMBAR PERSETUJUAN PENGUJI

### LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010034  
Nama : Ahmad Syauqi Irhamna  
Judul Tugas Akhir : Penerapan Algoritma Naïve Bayes Classifier  
Dalam Klasifikasi Curah Hujan DKI Jakarta

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 27 Juli 2022



U N I (Dr. Harwikarya, M.T)

MERCU BUANA



## LEMBAR PENGESAHAN

### LEMBAR PENGESAHAN

NIM : 41518010034  
Nama : Ahmad Syauqi Irfhamna  
Judul Tugas Akhir : Penerapan Algoritma Naïve Bayes Classifier Dalam  
Klasifikasi Curah Hujan DKI Jakarta

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 27 Juli 2022

Menyetujui,



(Yaya Sudarya Triana, Ph.D.)

Dosen Pembimbing

UNIVERSITAS  
MERCU BUANA

Mengetahui,



(Wawan Gunawan, S.Kom, MT)

Koord. Tugas Akhir Teknik Informatika



(Ir. Emu R. Kaburuan, Ph.D., I.F.M.)

Ka. Prodi Teknik Informatika

## KATA PENGANTAR

Segala puji dan syukur kehadiran Allah SWT atas berkah, rahmat dan hidayah-nya yang senantiasa dilimpahkan kepada penulis, sehingga bisa menyelesaikan skripsi dengan judul “Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Curah Hujan DKI Jakarta” sebagai syarat untuk menyelesaikan Program Sarjana (S1) pada Program Sarjana Fakultas Ilmu Komputer Jurusan Informatika Universitas Mercu Buana.

Dalam penyusunan skripsi ini banyak hambatan serta rintangan yang penulis hadapi namun pada akhirnya dapat melaluinya berkat adanya bimbingan dan bantuan dari berbagai pihak baik secara moral maupu spiritual. Untuk itu pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada:

1. Kedua orang tua yang telah memberika support serta kasih sayang hingga saya bisa menyelesaikan tugas akhir ini
2. Bapak Yaya Sudarya Triana, Ph.D. selaku dosen pembimbing saya yang telah membantu dan memberikan arahan kepada saya dalam menyelesaikan tugas akhir ini.
3. Bapak Emil R. Kaburuan, Ph.D. selaku Kepala Program Studi Informatika yang telah memberikan arahan dan bimbingannya selama perkuliahan.
4. Bapak Wawan Gunawan S.Kom, MT selaku Koordinator Tugas Akhir Prodi Informatika yang telah memberikan arahan selama periode tugas akhir
5. Bapak Dr. Harwikarya, MT selaku dosen pembimbing akademik saya yang telah membimbing selama perkuliahan hingga sampai pada tugas akhir.
6. Teman-teman dan semua pihak yang selalu memberikan support dan waktu untuk selalu mengingatkan dalam menyelesaikan tugas akhir ini.

Penulis menyadari bahwa penyusunan Jurnal Tugas Akhir ini masih jauh dari kesempurnaan dan masih terdapat kekurangan dan kelemahan, walaupun demikian, penulis telah berusaha semaksimal mungkin untuk mendapatkan hasil yang terbaik. Akhir kata, penulis berharap jika ada kritik atau saran apapun yang sifatnya membangun bagi saya dengan senang hati akan saya terima.

Jakarta, 07 Juli 2022

Ahmad Syauqi Irhamna

xi

## DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL .....	i
LEMBAR PERNYATAAN ORISINALITAS .....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI .....	v
LEMBAR PENGESAHAN .....	viii
ABSTRAK .....	ix
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xii
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR.....	xiv
NASKAH JURNAL .....	1
KERTAS KERJA.....	7
BAB 1. LITERATUR REVIEW .....	8
BAB 2. ANALISIS DAN PERANCANGAN.....	13
BAB 3. SOURCE CODE .....	15
BAB 4. DATASET.....	17
BAB 5. TAHAPAN EKSPERIMEN.....	21
BAB 6. HASIL SEMUA EKSPERIMEN.....	29
DAFTAR PUSTAKA .....	46
LAMPIRAN DOKUMEN HAKI.....	49
LAMPIRAN KORESPONDENSI .....	51

## DAFTAR TABEL

Tabel 1 Literatur Review Jurnal.....	9
Tabel 2 Variabel Penelitian.....	18
Tabel 3 Kategori Variabel Dependen.....	19
Tabel 4 Variabel Independen .....	20
Tabel 5 Confussion Matrix .....	27
Tabel 6 Penggunaan Data Training dan Data Testing .....	33
Tabel 7 Priori Probability Data Training .....	33
Tabel 8 Hasil Confussion Matrix NBC Data Training.....	34
Tabel 9 Priori Probability Data Testing .....	35
Tabel 10 Hasil Confussion Matrix NBC Data Testing .....	36



## DAFTAR GAMBAR

Gambar 1 Flowchart Penelitian.....	21
Gambar 2 Data Curah Hujan Harian Jakarta.....	22
Gambar 3 Data Cleaning Curah Hujan Harian DKI Jakarta .....	22
Gambar 4 Tahap Klasifikasi Penelitian.....	23
Gambar 5 Jumlah Status Hujan DKI Jakarta Tahun 2017 – 2021 .....	29
Gambar 6 Jumlah Status Hujan Ekstrem DKI Jakarta Tahun 2017 – 2021.....	30
Gambar 7 Jumlah Status Tidak Hujan DKI Jakarta Tahun 2017 – 2021.....	30
Gambar 8 Deskriptif Variabel Lama Penyinaran dan Kecepatan Angin .....	31
Gambar 9 Deskriptif Variabel Suhu dan Kelembaban.....	32
Gambar 10 Hasil Akurasi dan Total Data Training .....	35
Gambar 11 Hasil Akurasi dan Total Data Testing .....	37
Gambar 12 Output Source Code Data Training.....	40
Gambar 13 Output Source Code Data Testing.....	43

UNIVERSITAS  
MERCU BUANA

# Penerapan Algoritma *Naïve Bayes Classifier* Dalam Klasifikasi Curah Hujan DKI Jakarta

Ahmad Syauqi Irhamna<sup>1</sup>, Yaya Sudarya Triana<sup>2\*)</sup>

<sup>1</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercubuana, Jakarta

<sup>2</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercubuana, Jakarta

<sup>1,2</sup>Jln. Raya Meruya Selatan, Kec Kembangan, Jakarta, Daerah Khusus Ibukota Jakarta 11650, Indonesia

email: <sup>1</sup>[asirhamna@gmail.com](mailto:asirhamna@gmail.com), <sup>2\*)</sup>[yaya.sudarya@mercubuana.ac.id](mailto:yaya.sudarya@mercubuana.ac.id)

**Abstract** – *Rainfall is one of the elements in the climate. High rainfall intensity or what is often called extreme rain can result in flooding. To minimize the occurrence of floods caused by rainfall, a daily rainfall forecast is needed. Rainfall forecasting has become one of the most scientific and technologically challenging problems around the world in the last century. The Naïve Bayes Classifier (NBC) method, a classification method that can be used for rainfall forecasting in this study. The data used came from the BMKG of all stations in DKI Jakarta province from 2017 to 2021 with four independent variables, namely Average Temperature (°C), Average Humidity (%), Irradiation Duration (hours), and Average Wind Speed (knots). The results of the study showed that the NBC method training data produced an accuracy of 0.96710 or 96.710%.*

**Abstrak** – Curah hujan termasuk salah satu elemen pada iklim. Intensitas curah hujan yang tinggi atau yang sering disebut hujan ekstrem dapat mengakibatkan terjadinya banjir. Untuk meminimalisir terjadinya banjir yang diakibatkan oleh curah hujan, maka diperlukan prakiraan curah hujan harian. Prakiraan curah hujan telah menjadi salah satu masalah yang paling ilmiah dan teknologi menantang seluruh dunia pada abad terakhir. Metode Naïve Bayes Classifier (NBC), metode klasifikasi yang dapat digunakan untuk prakiraan curah hujan pada penelitian ini. Data yang digunakan berasal dari BMKG seluruh stasiun di provinsi DKI Jakarta dari tahun 2017 hingga tahun 2021 dengan empat variabel independen yaitu Rata-rata Suhu (°C), Rata-rata Kelembaban (%), Lama Penyinaran (jam), dan Rata-rata Kecepatan Angin (knot). Hasil dari penelitian menunjukkan untuk data training metode NBC menghasilkan akurasi sebesar 0,96710 atau sebesar 96,710%.

**Kata Kunci** – *Curah Hujan, Klasifikasi Naïve Bayes Classifier.*

## I. PENDAHULUAN

Banjir di Indonesia khususnya di DKI Jakarta diperkirakan terjadi hampir setiap tahun, hal ini disebabkan tingginya intensitas hujan. Curah hujan juga merupakan faktor yang secara langsung mempengaruhi perubahan cuaca, baik cuaca yang menguntungkan maupun cuaca yang buruk. Beberapa faktor kunci yang dapat mempengaruhi terjadinya presipitasi antara lain suhu udara, kelembaban udara, kecepatan angin, dan waktu penyinaran [1]. Iklim merupakan bagian penting dari geografi, karena iklim memiliki pengaruh yang besar terhadap aktivitas manusia di

lingkungan, seperti transportasi, pertanian, perkebunan dan lain-lain. Selain curah hujan, cuaca juga menjadi faktor penyebab banjir. Tidak seperti iklim, aktivitas atmosfer dalam jangka waktu yang lama dengan kondisi akumulatif kondisi cuaca rata-rata seperti angin, suhu dan curah hujan[2].

Kebutuhan prakiraan cuaca yang akurat juga diharapkan akan sangat bermanfaat bagi masyarakat dalam aktivitas sehari-hari yang membutuhkan informasi lebih akurat terkait prakiraan cuaca. Badan Meteorologi dan Geofisika Federal (BMKG) mengukur faktor cuaca setiap hari. Ketersediaan data yang melimpah pada faktor meteorologi ini mendukung ekstraksi informasi untuk prakiraan cuaca menggunakan data mining[3]. Data Mining merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola dan hubungan dalam set data ukuran besar[4][5]. Data mining dibagi menjadi enam kelompok berdasarkan tugas yang dapat dilakukan, yakni deskripsi, klasifikasi, prediksi, estimasi, kluster, dan asosiasi. Teknik prediksi (forecasting) melakukan estimasi terhadap sebuah data baru. Teknik ini mirip dengan klasifikasi, tetapi variabel target lebih kearah numerik dari pada kategorik. Biasanya jenis data pada teknik ini adalah runtun waktu (time series) dalam rentang waktu dari jam, harian, mingguan, bulanan bahkan tahunan. Masalah yang diprediksipun bervariasi, seperti perkiraan cuaca, tingkat inflasi, transaksi retail, maupun kurs mata uang[6].

Klasifikasi membantu membedakan kelas yang diketahui untuk tujuan memprediksi yang tidak diketahui. Salah satu metode klasifikasi yang sering digunakan adalah Naïve Bayes Classifier[7]. *Naïve bayes Classifier* merupakan pengklasifikasian sederhana yang menghitung probabilitas dengan menjumlahkan frekuensi dan nilai dari data yang ada. Naïve Bayes berdasar kepada teorema bayes yang digunakan untuk menghitung probabilitas dari tiap kelas dengan asumsi kelas satu dengan yang lain independen (tidak saling tergantung)[8].

Pada penelitian ini, penulis melakukan klasifikasi menggunakan metode Naïve Bayes Classifier (NBC), metode ini bertujuan untuk mencari status curah hujan.

\*) penulis korespondensi: Yaya Sudarya Triana

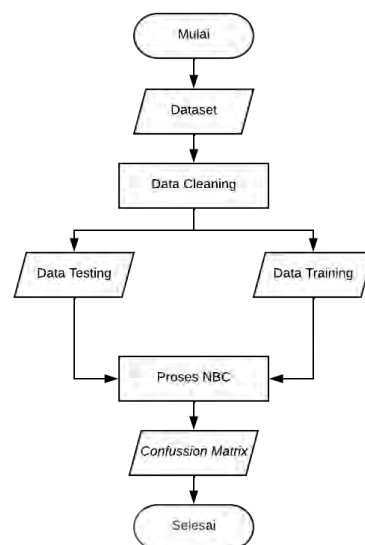
Email: [yaya.sudarya@mercubuana.ac.id](mailto:yaya.sudarya@mercubuana.ac.id)

## II. PENELITIAN YANG TERKAIT

Penelitian yang melakukan prediksi status curah hujan dengan metode naïve bayes telah banyak yang melakukan. Sedangkan penelitian ini mengusulkan status curah hujan menjadi tiga status curah hujan yaitu hujan, hujan ekstrem, dan tidak hujan, dengan data hasil pengumpulan melalui seluruh stasiun BMKG di DKI Jakarta. Misalnya, penelitian terkait perbandingan klasifikasi curah hujan dengan metode SVM dan NBC, data yang digunakan tahun 2018 di DKI Jakarta. Dari hasil analisis klasifikasi yang didapat bahwa metode terbaik yaitu Support Vector Machine hal ini dibuktikan dengan tingkat akurasi sebesar 79,45 % lebih besar dari tingkat akurasi metode Naïve Bayes Classifier yaitu 65,75%[1]. Memprediksi curah hujan di provinsi Bali menggunakan metode Naïve Bayesian. Penelitian ini dengan pengujian terdiri dari 132 data dengan masing-masing kategori hujan sebanyak 88 data. Hasil akhir pengujian ini didapatkan nilai akurasi untuk klasifikasi hujan dengan kategori ringan, normal, dan deras berturut turut 79.5%, 40.5% dan 86.4%[9]. Memprediksi status gunung berapi menggunakan dua algoritma klasifikasi KNN dan Naïve Bayes Classifier dengan metode K-Fold Cross. Penelitian ini bertujuan untuk membandingkan dua algoritma untuk validasi data untuk prediksi aktivasi status gunung berapi. Hasil akhir penelitian yang dibagi kedalam 3 fold di setiap metode klasifikasi diperoleh perbandingan akurasi sistem rata-rata tertinggi pada k-nn 63,68 % dengan standar deviasi 7,47 %. Sedangkan metode naive bayes diperoleh rata-rata akurasi sebesar 79,71 % dengan standar deviasi 3,55 %. Selain itu, penggunaan *naive bayes* jaraknya akurasi lebih dekat dibandingkan dengan k-nn[10].

## III. METODE PENELITIAN

Penelitian ini bertujuan untuk memprediksi curah hujan yang dikombinasikan dengan klasifikasi algoritma *naive bayes classifier*. Tahapan awal dilakukan dengan pengumpulan dataset dari BMKG. Tahap berikutnya melakukan cleaning data untuk menghilangkan data missing dan variabel yang tidak digunakan. Berikutnya membagi jenis data menjadi data training dan data testing. Tahapan selanjutnya adalah melakukan proses pelatihan dataset tersebut menggunakan metode *naive bayes classifier*. Tahapan selanjutnya setelah diperoleh model dari model terbaik pada data *training*, maka akan dilakukan proses klasifikasi dengan data *testing*. Selanjutnya setelah proses untuk data training dilakukan, maka akan menghasilkan tabel confusion matrix untuk data training. Pada data testing juga dilakukan proses *naive bayes classifier* dari model yang telah diperoleh dengan data training, sehingga data testing akan menghasilkan confusion matrix untuk digunakan digunakan untuk mencari nilai akurasi pada metode *naive bayes classifier*.



Gbr. 1 Ilustrasi metode penelitian yang digunakan.

### A. Dataset

Penelitian ini menggunakan data yang bersumber dari website resmi BMKG dataset yang digunakan adalah curah hujan harian di Provinsi DKI Jakarta mulai 01 Januari 2017 hingga 31 Desember 2021 pada seluruh stasiun BMKG di DKI Jakarta.

TABEL I  
DATASET

Jenis Variabel	Nama Variabel	Skala
Variabel Independen (X)	Rata-rata Suhu (°C)	Rasio
	Rata-rata Kelembaban (%)	Nominal
	Rata-rata Lama Penyinaran (jam)	Nominal
	Rata-rata Kecepatan Angin (knot)	Rasio
Variabel Dependen (Y)	Status	Ordinal

TABEL II  
KATEGORI VARIABEL DEPENDEN

Variabel	Definisi Operasional Variabel		
	Hujan	Hujan Ekstrem	Tidak Hujan
Status	Jika nilai pada variabel curah hujan $> 0$ dan $\leq 50$ mm / hari.	Jika nilai pada variabel curah hujan $> 50$ mm / hari	Jika nilai pada curah hujan = 0 mm / hari.



### B. Data Cleaning

Tahap kedua pada penelitian ini adalah proses pembersihan data bertujuan agar dataset tidak mengandung *noise* yang masih *outliers* atau *error*, data *incomplete* merupakan data yang nilai atributnya hilang, dan data *inconsistent* data yang tidak konsisten sehingga dapat mempengaruhi hasil klasifikasi.

### C. Data Training dan Data Testing

Data yang telah dipraproses dibagi menjadi data latih dan data validasi. Data latih digunakan untuk membangun model Artificial Neural Network, Decision tree, dan Naïve Bayes melalui proses pelatihan, sedangkan data validasi digunakan untuk memvalidasi model yang telah dibangun pada proses pelatihan sebelumnya. Data yang digunakan untuk training adalah 80% dan sisanya sebesar 20% digunakan untuk testing [11].

### D. Proses Naïve Bayes Classifier

Naïve Bayes Classifier (NBC) menggunakan teori probabilitas sebagai dasar teori. Dalam bukunya, Han, J. dan Kamber, M. menyatakan bahwa "Bayesian classifiers mempunyai tingkat kecepatan dan akurasi yang tinggi ketika diaplikasikan dalam database yang besar" [12]. Naïve Bayes Classifier merupakan pengklasifikasian sederhana yang menghitung probabilitas dengan menjumlahkan frekuensi dan nilai dari data yang ada. Naïve Bayes berdasar kepada teorema Bayes yang digunakan untuk menghitung probabilitas dari tiap kelas dengan asumsi kelas satu dengan yang lain independen (tidak saling tergantung) [8]. Rumus NBC untuk klasifikasi adalah sebagai berikut:

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X_i)} \quad (1)$$

Setiap  $X = \{X_1, X_2, X_3, \dots, X_q\}$  sebanyak  $q$  atribut atau  $q$  dimensi dimana :

$P(Y|X)$  = Probabilitas data dengan vektor  $X$  pada kelas  $Y$

$P(Y)$  = Probabilitas awal kelas  $Y$  (*Priori probability*)

$P(X|Y)$  = Probabilitas akhir (*Posterior Probability*)

$\prod_{i=1}^q P(X_i|Y)$  = Probabilitas independen dari kelas  $Y$  dalam vektor  $X$

$i$  = Observasi ke- $i$

### E. Confusion Matrix

Aktual merupakan klasifikasi status hujan yang sebelumnya telah diklasifikasikan terlebih dahulu.

Prediksi merupakan hasil dari klasifikasi variabel status yang dihasilkan oleh program.

Beberapa persyaratan yang telah didefinisikan untuk matrik klasifikasi di antaranya sebagai berikut:

$$\text{Akurasi} = \frac{F_{11}+F_{22}}{F_{11}+F_{21}+F_{12}+F_{22}} \quad (2)$$

$$\text{True Positive} = \frac{F_{22}}{F_{12}+F_{22}} \quad (3)$$

$$\text{False Positive} = \frac{F_{11}+F_{21}}{F_{11}} \quad (4)$$

$$\text{True Negative} = \frac{F_{11}}{F_{11}+F_{21}} \quad (5)$$

$$\text{False Negative} = \frac{F_{21}}{F_{12}+F_{22}} \quad (6)$$

$$P = \frac{F_{22}}{F_{21}+F_{22}} \quad (7)$$

Keterangan :

Accuracy : proporsi jumlah prediksi benar.

Recall atau true positive (TP) : proporsi kasus positif yang diidentifikasi dengan benar

Tingkat false positive (FP) : proporsi kasus negatif yang salah diklasifikasikan sebagai positif

Tingkat true negative (TN) : proporsi kasus negatif yang diklasifikasikan dengan benar

Tingkat false negative (FN) : proporsi kasus positif yang salah diklasifikasikan sebagai negatif

Precision (P) : proporsi prediksi kasus positif yang benar

## IV. HASIL DAN PEMBAHASAN

### A. Pembagian Data Training dan Data Testing

Data pada klasifikasi umumnya dibagi menjadi dua bagian yaitu data training dan data testing.

TABEL III  
PENGUNAAN DATA TRAINING DAN DATA TESTING

Data	Jumlah
Training	2.280
Testing	570
Total	2.850

Berdasarkan tabel III, total data yang digunakan pada penelitian ini adalah 2.850, dengan komposisi 80% untuk data training yaitu sebesar 2.280 data, dan sisanya 570 untuk data testing. Jumlah data training sebesar 80% tersebut ditentukan oleh peneliti berdasarkan penelitian sebelumnya.

### B. Analisis Naïve Bayes Classifier

Implementasi penelitian ini dilakukan dengan menggunakan metode NBC untuk menentukan nilai peluang dalam menentukan kelas klasifikasi. Kelas klasifikasi yang dihasilkan adalah hujan, hujan ekstrem,

Universitas Mercu Buana

dan tidak hujan dengan jumlah data 2.850 akan menghasilkan kelas klasifikasi dan tabel prediksi. Hasil *priori probability* data training dijelaskan pada Tabel IV.

TABEL IV  
PRIORI PROBABILITY DATA TRAINING

Hujan	Hujan Ekstrem	Tidak Hujan
0.46535088	0.06578947	0.46885965

Berdasarkan data tersebut dapat disimpulkan bahwa probabilitas hujan yaitu sebesar 0.46535 diperoleh dari 1.061 dibagi dengan 2.280. Probabilitas prediksi hujan ekstrem sebesar 0.06578 diperoleh dari 150 dibagi dengan 2.280. Untuk probabilitas prediksi status tidak hujan sebesar 0.46885 diperoleh dari 1.069 dibagi dengan 2.280. Selain itu total prediksi masing-masing status hujan bisa di jelaskan pada Tabel V.

TABEL V  
HASIL CONFUSION MATRIX NBC DATA TRAINING

Observasi	Prediksi			Total
	Hujan	Hujan Ekstrem	Tidak Hujan	
Hujan	992	0	0	992
Hujan Ekstrem	6	150	0	156
Tidak Hujan	63	0	1.069	1.132
Total	1.061	150	1.069	2.280

TABEL VI  
PRIORI PROBABILITY DATA TESTING

Hujan	Hujan Ekstrem	Tidak Hujan
0.43333333	0.06842105	0.49824561

Berdasarkan Tabel VI diatas diperoleh nilai probabilitas awal untuk masing-masing status pada variabel status hujan, hujan ekstrem, dan tidak hujan. Total untuk prediksi masing-masing status hujan bisa dilihat pada tabel 6.3. Untuk probabilitas hujan yaitu sebesar 0.43333 diperoleh dari 247 dibagi dengan 570. Probabilitas prediksi hujan ekstrem sebesar 0.06842 diperoleh dari 39 dibagi dengan 570. Untuk probabilitas prediksi status tidak hujan sebesar 0.49824 diperoleh dari 284 dibagi dengan 570.

TABEL VII  
HASIL CONFUSION MATRIX NBC DATA TESTING

Observasi	Prediksi			Total
	Hujan	Hujan Ekstrem	Tidak Hujan	
Hujan	243	0	0	243

Hujan Ekstrem	0	39	0	39
Tidak Hujan	4	0	284	288
Total	247	39	284	570

Melihat hasil perbandingan data antar observasi dan prediksi pada Tabel V dan Tabel VII. Hasil untuk data training diperoleh nilai akurasi sebesar 0,96973 atau sebesar 96,973%. Hasil akurasi tersebut merupakan hasil akurasi yang cukup besar, sedangkan untuk nilai error sebesar 0,03026, sedangkan hasil untuk data testing diperoleh nilai akurasi sebesar 0,99298 atau 99,298%. Untuk hasil error data testing diperoleh nilai sebesar 0,00701.

### C. Hasil Model dan Plot

```

-----
::: Rata.rata.kelembapan....::Hujan (KDE)
-----

```

```

Call:
  density.default(x = x, na.rm = TRUE)

Data: x (1061 obs.);   Bandwidth 'bw' = 1

      x              y
Min.  :62.00   Min.  :8.910e-06
1st Qu.:70.75   1st Qu.:2.504e-03
Median :79.50   Median :1.394e-02
Mean   :79.50   Mean   :2.854e-02
3rd Qu.:88.25   3rd Qu.:5.284e-02
Max.   :97.00   Max.   :8.619e-02

```

```

-----
::: Rata.rata.kelembapan....::HujanEkstrem (KDE)
-----

```

```

Call:
  density.default(x = x, na.rm = TRUE)

Data: x (150 obs.);   Bandwidth 'bw' = 2.138

      x              y
Min.  : 50.58   Min.  :1.404e-05
1st Qu.: 63.29   1st Qu.:1.344e-03
Median : 76.00   Median :1.018e-02
Mean   : 76.00   Mean   :1.965e-02
3rd Qu.: 88.71   3rd Qu.:3.512e-02
Max.   :101.42   Max.   :6.884e-02

```

```
---: Rata.rata.Kelembapan....:Tidak Hujan (KDE)
```

Call:

```
density.default(x = x, na.rm = TRUE)
```

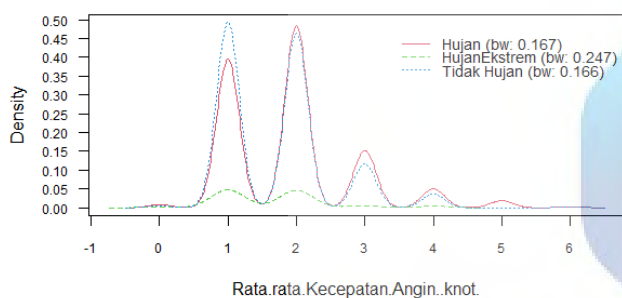
Data: x (1069 obs.); Bandwidth 'bw' = 0.9988

x		y	
Min.	:51	Min.	:4.260e-06
1st Qu.	:62	1st Qu.	:4.285e-04
Median	:73	Median	:9.669e-03
Mean	:73	Mean	:2.271e-02
3rd Qu.	:84	3rd Qu.	:3.967e-02
Max.	:95	Max.	:7.702e-02

Gbr. 1 Hasil Model Data Training

\$Matrix\_Uji

pred22	Hujan	HujanEkstrem	Tidak Hujan
Hujan	992	0	0
HujanEkstrem	6	150	0
Tidak Hujan	63	0	1069



Gbr. 2 Akurasi dan Plot Data Training

```
---: Rata.rata.Kelembapan....:Hujan (KDE)
```

Call:

```
density.default(x = x, na.rm = TRUE)
```

Data: x (247 obs.); Bandwidth 'bw' = 1.116

x		y	
Min.	:64.65	Min.	:1.638e-05
1st Qu.	:72.83	1st Qu.	:2.930e-03
Median	:81.00	Median	:1.715e-02
Mean	:81.00	Mean	:3.056e-02
3rd Qu.	:89.17	3rd Qu.	:5.008e-02
Max.	:97.35	Max.	:1.079e-01

```
---: Rata.rata.Kelembapan....:HujanEkstrem (KDE)
```

Call:

```
density.default(x = x, na.rm = TRUE)
```

Data: x (39 obs.); Bandwidth 'bw' = 1.937

x		y	
Min.	:61.19	Min.	:6.091e-05
1st Qu.	:69.84	1st Qu.	:6.657e-03
Median	:78.50	Median	:1.872e-02
Mean	:78.50	Mean	:2.885e-02
3rd Qu.	:87.16	3rd Qu.	:4.617e-02
Max.	:95.81	Max.	:8.549e-02

```
---: Rata.rata.Kelembapan....:Tidak Hujan (KDE)
```

Call:

```
density.default(x = x, na.rm = TRUE)
```

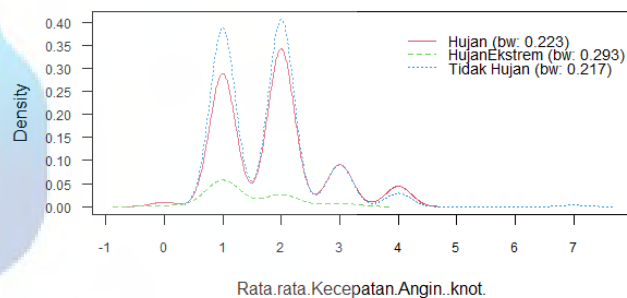
Data: x (284 obs.); Bandwidth 'bw' = 1.302

x		y	
Min.	:54.09	Min.	:1.219e-05
1st Qu.	:63.55	1st Qu.	:1.606e-03
Median	:73.00	Median	:1.464e-02
Mean	:73.00	Mean	:2.642e-02
3rd Qu.	:82.45	3rd Qu.	:4.977e-02
Max.	:91.91	Max.	:8.222e-02

Gbr. 3 Hasil Model Data Testing

\$Matrix\_Uji

predtest	Hujan	HujanEkstrem	Tidak Hujan
Hujan	243	0	0
HujanEkstrem	0	39	0
Tidak Hujan	4	0	284



Gbr. 4 Akurasi dan Plot Data Testing

UNIVERSITAS

MERCU BUANA

## V. KESIMPULAN

Berdasarkan analisis pembahasan diperoleh hasil akurasi untuk data training dan data testing dari metode Naive Bayes Classifier dengan data training sebesar diperoleh nilai akurasi 0,96710 atau sebesar 96,710%. Hasil akurasi tersebut merupakan hasil akurasi yang cukup besar, sedangkan untuk nilai error sebesar 0,03289, sedangkan hasil untuk data testing diperoleh nilai akurasi sebesar 0,99298 atau 99,298%. Untuk hasil error data testing diperoleh nilai sebesar 0,00701.

## DAFTAR PUSTAKA

- [1] M. L. Laia and Y. Setyawan, "Perbandingan Hasil Klasifikasi Curah Hujan Menggunakan Metode SVM dan NBC," vol. 05, no. 2, pp. 51–61, 2020.
- [2] D. D. Sidik and T. W. Sen, "Penggunaan Stacking Classifier Untuk Prediksi Curah Hujan," *IT Soc.*, vol. 4, no. 1, pp. 21–27, 2019, doi: 10.33021/itfs.v4i1.1180.
- [3] B. P. T.P and R. D. Indah Sari, "Penerapan Data Mining Untuk Prakiraan Cuaca Di Kota Malang Menggunakan Algoritma Iterative Dichotomiser Tree (Id3)," *Jouticla*, vol. 2, no. 2, pp. 101–108, 2017, doi: 10.30736/jti.v2i2.68.
- [4] A. Nurlifa and S. Kusumadewi, "Sistem Peramalan Jumlah

Universitas Mercu Buana

- Penjualan Menggunakan Metode Moving Average Pada Rumah Jilbab Zaky,” *INOVTEK Polbeng - Seri Inform.*, vol. 2, no. 1, p. 18, 2017, doi: 10.35314/isi.v2i1.112.
- [5] Iwan, E. R. Iviq, and A. Yulianto, “Analisa Peramalan Permintaan Mobil Mitsubishi Xpander dengan Tiga Metode Forecasting,” *J. Hum.*, vol. 18, no. 2, pp. 249–256, 2018, [Online]. Available: doi: <https://doi.org/10.31294/jc.v18i2>.
- [6] A. D. W. Sumari, M. B. Musthafa, Ngatmari, and D. R. H. Putra, “Perbandingan Kinerja Metode-Metode Prediksi pada Transaksi Dompot Digital di Masa Pandemi,” *Rekayasa Sist. dan Teknol. Inf.*, vol. 4, no. 4, pp. 642–647, 2017.
- [7] H. Muhamad, C. A. Prasojo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, “Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 180, 2017, doi: 10.25126/jtiik.201743251.
- [8] E. Sutoyo and A. Almaarif, “Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 95–101, 2020, doi: 10.29207/RESTI.V4I1.1502.
- [9] I. G. A. Gunadi, A. Aprilyana, and K. Dewi, “Klasifikasi Curah Hujan di Provinsi Bali Berdasarkan Metode Naïve Bayesian,” *Wahana Mat. dan Sains J. Mat. Sains, dan Pembelajarannya*, vol. 12, no. 1, pp. 14–25, 2018.
- [10] F. Tempola, M. Muhammad, and A. Khairan, “Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 577, 2018, doi: 10.25126/jtiik.201855983.
- [11] S. Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, N. Chamidah, M. Mega Santoni, N. Matondang, F. Ilmu Komputer, and U. Pembangunan Nasional Veteran Jakarta, “Terakreditasi SINTA Peringkat 2 Pengaruh Oversampling pada Klasifikasi Hipertensi dengan Algoritma Naïve Bayes, Decision Tree, dan Artificial Neural Network (ANN),” *Masa Berlaku Mulai*, vol. 1, no. 3, pp. 635–641, 2017.
- [12] R. A. Dwi Pramita B. B1, Ristu Saptono2, “Academic Articles Classification Using Naive Bayes Classifier (Nbc) Method,” *J. Ilm. Teknol. dan Inf.*, vol. 7, no. 2, pp. 74–81, 2018.



## KERTAS KERJA

Kertas kerja ini merupakan materi lengkap dari artikel jurnal yang berjudul "Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Curah Hujan DKI Jakarta". Kertas kerja ini berisi semua materi yang diperoleh dari survei tugas akhir. Di kertas kerja ini berisi beberapa bagian yang terdiri dari literature review, analisis dan perancangan, dataset yang digunakan, tahapan eksperimen, dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan mengenai analisis masalah yang terjadi. Bagian III menjelaskan tentang source code yang digunakan dalam penelitian ini. Bagian IV menjelaskan mengenai dataset yang digunakan, termasuk penjelasan dan sumber data. Bagian V berisi tahapan eksperimen yang ditunjukkan pada gambar dengan penjelasan dari setiap tahapan. Bagian VI merupakan bagian akhir dari kertas kerja ini dan menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, termasuk penjelasannya.



UNIVERSITAS  
MERCU BUANA