

Penerapan Algoritma Gaussian Naive Bayes Dalam Klasifikasi Cuaca Kota Tanjungpinang

Ganda Bagus Wibisono^{1,*}, Martaleli Bettiza^{2,*}, Nurfalinda³

^{1,2,3}Jurusan Teknik Informatika, Fakultas Teknik dan Teknik Kemaritiman,
Universitas Maritim Raja Ali Haji

^{1,2,3}Jl. Politeknik Senggarang, Tanjungpinang 29100

*Corresponding Author: mbettiza@umrah.ac.id

Abstract—Weather refers to the instantaneous atmospheric conditions at a specific place and time. Weather can change within a few hours. Weather consists of several weather elements. Accurate weather forecasts can help the daily activities of citizens obtained from proper classification. The foundation of Bayes classification is the theorem of conditional probability. Naïve Bayes algorithm and Gaussian Naïve Bayes algorithm are different from each other. The Gaussian Naïve Bayes technique incorporates a Gaussian distribution to overcome these differences. Raja Haji Fisabilillah Tanjungpinang Meteorological Station provides data on temperature, humidity, air pressure, and wind speed used as research parameters. The data is daily data covering the period January 1, 2019 to December 31, 2019 with a total of 365 data. The research results are categorized into several types of weather, namely thunderstorm (TS), Lightning, Mist, lightning rain (TS/RA), Cloudy, rain (RA) and haze (HZ). The output of this study obtained the highest accuracy of 60.87% with 50% training data.

Keywords—Weather, Naïve Bayes, Gaussian, Classification.

Intisari—Cuaca mengacu pada kondisi atmosfer sesaat di tempat dan saat tertentu. Cuaca dapat berubah dalam beberapa jam. Cuaca terdiri dari beberapa elemen cuaca. Perkiraan cuaca yang akurat bisa membantu kegiatan sehari-hari warga yang didapat dari klasifikasi yang tepat. Landasan dari klasifikasi Bayes adalah teorema tentang probabilitas bersyarat. Algoritma Naïve Bayes dan algoritma Gaussian Naïve Bayes berbeda satu sama lain. Teknik ini menggabungkan distribusi Gaussian untuk mengatasi perbedaan tersebut. Stasiun Meteorologi Raja Haji Fisabilillah Tanjungpinang menyediakan data suhu, kelembaban, tekanan udara, dan kecepatan angin yang digunakan sebagai parameter penelitian. Data tersebut merupakan data harian yang mencakup periode 1 Januari 2019 hingga 31 Desember 2019 dengan total 365 data. Hasil penelitian dikategorikan ke dalam beberapa kategori cuaca yaitu thunderstorm (TS), Lightning, Mist, lightning rain (TS/RA), Cloudy, rain (RA) dan haze (HZ). Output dari penelitian ini didapatkan akurasi tertinggi sebesar 60,87% dengan jumlah data training sebanyak 50%.

Kata kunci—Cuaca, Naïve Bayes, Gaussian, Klasifikasi.

I. PENDAHULUAN

Cuaca mengacu pada kondisi atmosfer tertentu pada lokasi dan waktu tertentu, yang dicirikan oleh wilayah geografis yang relatif terbatas dan durasi yang singkat [1]. Kondisi cuaca yang berubah dengan cepat yang

diakibatkan oleh konvergensi berbagai elemen atmosfer dan pola cuaca dalam rentang waktu yang singkat dalam beberapa jam. Cuaca dan iklim dipengaruhi oleh berbagai faktor, termasuk suhu udara, tekanan udara, kelembaban udara, dan curah hujan. dan curah hujan [2].

Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) di Indonesia menyediakan prakiraan cuaca yang menilai kondisi meteorologi sepanjang hari. BMKG adalah badan pemerintah yang bertanggung jawab untuk memantau dan memprediksi fenomena meteorologi. Data cuaca disebarluaskan kepada masyarakat umum setiap hari untuk melayani berbagai tujuan, termasuk transportasi, kesehatan, pertanian, perencanaan infrastruktur, pariwisata, dan banyak lagi.

Klasifikasi yang tepat akan sangat bermanfaat bagi individu dalam usaha sehari-hari yang membutuhkan prediksi meteorologi yang lebih akurat. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) melakukan pengukuran parameter meteorologi dan klimatologi setiap hari. Data mining dapat secara efektif mengekstrak informasi prakiraan cuaca dengan memanfaatkan kekayaan data faktor meteorologi yang tersedia. Data mining adalah proses pengumpulan dan pemanfaatan data masa lalu untuk mengidentifikasi pola, tren, dan hubungan yang konsisten di seluruh kumpulan data yang luas. Data mining dikategorikan ke dalam enam divisi yang berbeda berdasarkan tugas-tugas spesifik yang diselesaikannya, yaitu deskripsi, klasifikasi, prediksi, estimasi, pengelompokan, dan asosiasi

Klasifikasi Bayes didasarkan pada penggunaan teorema Bayes, sebuah rumus matematika yang menghitung probabilitas bersyarat. Algoritme ini menghitung probabilitas bersyarat dari sebuah variabel stokastik (variabel kelas) berdasarkan pengamatan sebelumnya terhadap nilai sekelompok variabel stokastik lainnya (variabel fitur). Teorema Bayes digunakan secara luas dalam domain probabilitas dan statistik. Algoritma Gaussian Naive Bayes dan algoritma Naive Bayes berbeda satu sama lain. Perbedaannya terletak pada fakta bahwa algoritma Gaussian Naive Bayes menggunakan distribusi Gaussian.

Pada tahun 2022, Mahmuda melakukan penelitian mengenai klasifikasi curah hujan di Kota Semarang menggunakan pendekatan Gaussian Naive Bayes [3]. Penelitian ini menggunakan pendekatan Gaussian Naive Bayes untuk menganalisis data curah hujan harian di

sekitar Kota Semarang, meliputi periode 1 Januari hingga 30 Juni 2021. Data tersebut diperoleh dari situs web BMKG. Jumlah data yang diperoleh sebanyak 156 data. Variabel yang dimasukkan dalam penelitian ini meliputi suhu rata-rata, kelembaban udara rata-rata, kecepatan angin rata-rata, dan lama penyinaran matahari rata-rata. Curah hujan dikategorikan ke dalam empat klasifikasi: tidak ada curah hujan, gerimis, curah hujan sedang, dan curah hujan lebat. Penelitian dan perdebatan telah menghasilkan kesimpulan bahwa ketika mengevaluasi data dengan jumlah sampel 31 menggunakan metode Gaussian Naive Bayes, tingkat akurasi ditentukan sebesar 64,52%.

Berdasarkan paparan tersebut, maka akan dilakukan penelitian mengenai penerapan algoritma Gaussian Naive Bayes dalam mengklasifikasi cuaca, kemudian untuk mengetahui tingkat akurasi dilakukan dengan membandingkan data hasil pengujian dengan data aktual.

II. METODE PENELITIAN

Teknik penelitian ini mengkaji pendekatan yang digunakan dalam membuat sistem yang digunakan untuk analisis dan implementasi algoritma. Tujuan dari penelitian ini adalah untuk memanfaatkan pendekatan Gaussian Naive Bayes untuk klasifikasi cuaca di Kota Tanjungpinang.

Dalam penelitian ini, penulis menggunakan data sekunder berupa data cuaca yang di dapatkan melalui Stasiun Meteorologi Raja Haji Fisabilillah. Data yang digunakan berupa data harian dari tanggal 1 Januari 2019 sampai 31 Desember 2019 yang berjumlah 365 data. Data terdiri dari 4 unsur cuaca dan cuaca. Unsur cuaca akan digunakan sebagai variabel dan cuaca akan digunakan sebagai kelas yang berjumlah 8 kelas yaitu cloudy, Haze (HZ), Mist, Thunderstorm (TS), Rain (RA), Lightning, Smoke dan Lightning Rain (TS/RA).

Dalam penelitian ini data akan dibagi menjadi 2 yaitu data training dan data testing. Data testing akan digunakan untuk menguji performa dari model yang akan digunakan. Teknik untuk mencari model terbaik dengan

mencari model dengan akurasi tertinggi sehingga digunakan beberapa jumlah data training

A. Klasifikasi

Klasifikasi adalah bentuk analisis data yang mengekstrak model yang menggambarkan kelas data. Pengklasifikasi, atau model klasifikasi, memprediksi label kategorikal (kelas). Klasifikasi merupakan teknik untuk memetakan kelas ke dalam data yang telah terklasifikasi [4]. Prediksi numerik memodelkan fungsi-fungsi bernilai kontinu. Klasifikasi dan prediksi numerik adalah dua jenis utama masalah prediksi [5].

B. Naïve Bayes

Klasifikasi Naïve Bayesian didasarkan pada teorema Bayes tentang probabilitas posterior. Metode ini salah satu metode klasifikasi yang sederhana [6]. Metode ini mengasumsikan independensi kelas- kondisi-bahwa efek dari nilai atribut pada kelas yang diberikan tidak bergantung pada nilai atribut lainnya [3]. Naïve Bayes Classifier (NBC), juga dikenal sebagai Klasifikasi Bayesian, adalah teknik klasifikasi yang digunakan untuk menghitung kemungkinan suatu objek termasuk dalam kelas tertentu [7]. NBC adalah teknik kategorisasi yang bergantung pada prinsip-prinsip probabilistik dasar. Teorema Bayes memiliki kemampuan kategorisasi yang sebanding dengan metodologi Decision Tree dan Neural Network. Oleh karena itu, NBC menunjukkan efektivitas, efisiensi, dan keandalan dalam mengelola kumpulan data yang luas dan dapat secara efektif menangani data yang tidak biasa. X adalah vektor input yang berisi data, sedangkan Y mewakili label kelas.

C. Gaussian Naïve Bayes

Naïve Bayes memperlakukan atribut diskrit dan numerik agak berbeda. Pada Gaussian Naïve Bayes biasanya digunakan untuk merepresentasikan probabilitas bersyarat dari atribut numerik pada sebuah kelas [8]. Untuk setiap atribut diskrit, dimodelkan dengan satu bilangan real antara 0 dan 1 yang mewakili probabilitas bahwa atribut X akan mengambil nilai tertentu x ketika kelasnya adalah c. Sebaliknya, setiap atribut numerik dimodelkan dengan beberapa distribusi probabilitas kontinu

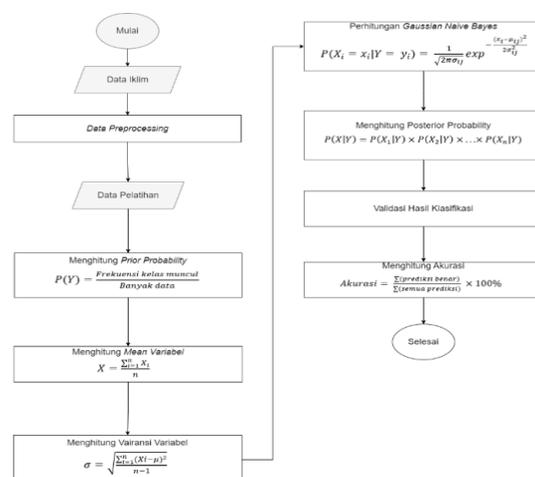
pada rentang nilai atribut tersebut [9]. Asumsi yang umum di dalam setiap kelas, nilai-nilai atribut numerik terdistribusi secara normal. Seseorang dapat merepresentasikan distribusi seperti itu dalam hal rata-rata dan standar deviasi dan seseorang dapat secara efisien menghitung probabilitas dari nilai yang diamati dari estimasi tersebut. Gaussian Naïve Bayes merupakan metode yang sederhana dalam penerapannya [10]. Untuk atribut kontinu dengan menggunakan persamaan distribusi Gaussian dinyatakan dengan persamaan berikut.

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma^2_{ij}}} \exp \left(-\frac{(x_i - \mu_{ij})^2}{2\sigma^2_{ij}} \right) \quad (1)$$

Dimana μ adalah rata-rata dari fitur dalam kelas Y, σ adalah deviasi standar, dan x adalah nilai fitur yang diamati. Integrasi konsep probabilitas dalam Gaussian Naïve Bayes membantu dalam peningkatan kinerja klasifikasi cuaca di wilayah Kota Tanjungpinang.

D. Flowchart Gaussian Naïve Bayes

Diagram alir menawarkan perspektif yang lebih komprehensif tentang tahapan berurutan yang terlibat dalam proses klasifikasi cuaca. Alur logika ini mencakup urutan tahapan penting dalam proses klasifikasi cuaca, dimulai dengan pengumpulan dan persiapan data dan berpuncak pada pencapaian hasil klasifikasi yang diperlukan. Diagram yang mengilustrasikan urutan langkah-langkah dalam perangkat lunak yang membandingkan metode Gaussian Naïve Bayes untuk klasifikasi cuaca disajikan pada Gambar 1.



Gambar 1. Flowchart Gaussian Naïve Bayes.

1. Data Preprocessing

Tahap pertama yang harus dilakukan adalah melakukan data selection untuk memilah data yang tidak diperlukan. Pada penelitian ini atribut yang diperlukan yaitu suhu, kelembapan, tekanan udara dan Kecepatan angin. Sedangkan kolom cuaca sebagai kelas.

2. Menghitung prior probability tiap kelas

$$P(Y) = \frac{\text{Frekuensi kelas muncul}}{\text{Banyak data}} \quad (2)$$

Dimana Y merupakan kelas.

3. Menghitung mean variable

$$X = \frac{\sum_{i=1}^n X_i}{n} \quad (3)$$

Dimana $\sum_{i=1}^n X_i$ jumlah data ke-I dan n merupakan banyak data.

4. Menghitung standar deviasi variabel

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n - 1} \quad (4)$$

Dimana Xi merupakan data ke-I, μ merupakan rata-rata variabel dan n merupakan banyak data.

5. Perhitungan distribusi Gaussian

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma^2_{ij}}} \exp \frac{(x_i - \mu_{ij})^2}{2\sigma^2_{ij}}$$

Dimana Y merupakan kelas, mean variabel, standar deviasi variabel.

6. Validasi hasil kualifikasi

Pada tahap ini dilakukan validasi dengan membandingkan data aktual dengan hasil klasifikasi data testing.

7. Menghitung Akurasi

Pada tahap ini dilakukan untuk melihat performa model dengan menghitung klasifikasi yang sama dengan data aktual.

$$\text{Akurasi} = \frac{\sum(\text{prediksi benar})}{\sum(\text{semua prediksi})} \times 100\% \quad (6)$$

III. HASIL DAN PEMBAHASAN

Bab ini akan menyajikan sistem manajemen data cuaca yang menangani data

harian dari tanggal 1 Januari 2019 hingga 31 Desember 2019. Data tersebut dicatat dalam basis data dan ditampilkan dalam sebuah tabel. Selanjutnya, sistem akan mengkategorikan data cuaca dan menentukan kebenaran algoritma Gaussian Naive Bayes. Data tersebut akan menjalani pengujian dengan menggunakan berbagai jumlah data pelatihan untuk mengidentifikasi model yang menunjukkan tingkat akurasi tertinggi.

A. Data Cuaca

Data cuaca yang digunakan berjumlah 365 data yang merupakan data cuaca per hari mulai tanggal 1 Januari 2019 sampai 31 Desember 2019. Data cuaca yang terdiri dari variabel tanggal, suhu, kelembapan, tekanan udara, kecepatan angin dengan jumlah kelas yang digunakan dibagi menjadi 8 yaitu cloudy, Haze (HZ), Mist, Thunderstorm (TS), Rain (RA), Lightning, Smoke dan Lightning Rain (TS/RA). dapat dilihat pada gambar 2.

Tanggal	Suhu	Kelembapan	Tekanan	Angin	Cuaca
2019-01-01	27.9	79	1010.6	08	TS
2019-01-02	28.9	73	1009.9	06	Lightning
2019-01-03	29.2	77	1009.5	07	Lightning
2019-01-04	29.3	77	1010.6	04	Mist
2019-01-05	27.0	93	1011.0	04	TS/RA
2019-01-06	28.0	86	1011.0	08	Cerah
2019-01-07	27.5	82	1010.0	10	Cerah
2019-01-08	28.0	82	1009.9	05	TS/RA
2019-01-09	28.0	81	1009.9	07	Cerah
2019-01-10	28.2	77	1010.2	05	Cerah
...
2019-12-31	27.7	82	1010.5	07	Cerah

Gambar 2. Data Cuaca 2019.

B. Data Hasil Pengujian

Berikut merupakan hasil pengujian klasifikasi cuaca menggunakan metode Gaussian Naive Bayes yang ditunjukkan pada tabel 3.

Tanggal	Suhu	Kelembapan	Tekanan	Kecepatan Angin	Cuaca	Hasil Testing
2019-10-19	28.0	83	1009.1	06	Lightning	Cerah
2019-10-20	28.0	85	1008.9	05	RA	RA
2019-10-21	28.0	83	1008.1	04	RA	RA
2019-10-22	26.5	88	1006.7	03	TS/RA	TS/RA
2019-10-23	26.9	87	1007.0	05	RA	TS/RA
...
2019-12-31	27.7	82	1010.5	07	Cerah	Cerah

Gambar 3. Data Hasil Pengujian.

Maka perhitungan persentase akurasi untuk data training 50% dilakukan dengan menggunakan persamaan 5.

$$\text{Persentase Akurasi} = \frac{112}{184} \times 100\% = 60,87\%$$

C. Analisis Metode Gaussian Naïve Bayes

Setelah dilakukan klasifikasi menggunakan metode Gaussian Naïve Bayes dengan beberapa jumlah data training, maka didapatkan model terbaik yaitu dengan data training sebanyak 50% dengan tingkat akurasi sebesar 60.87%. Berdasarkan hasil tersebut didapatkan bahwa metode Gaussian Naïve dalam klasifikasi cuaca Kota Tanjungpinang kurang baik dalam mengklasifikasikan cuaca di Kota Tanjungpinang.

IV. KESIMPULAN

Temuan penelitian menunjukkan bahwa algoritma Gaussian Naïve Bayes digunakan untuk klasifikasi cuaca di wilayah Kota Tanjungpinang. Dataset terdiri dari 365 data cuaca harian, yang dikategorikan ke dalam 8 jenis cuaca cloudy, Haze (HZ), Mist, Thunderstorm (TS), Rain (RA), Lightning, Smoke dan Lightning Rain (TS/RA). Elemen cuaca yang digunakan sebagai variabel untuk perhitungan adalah suhu rata-rata, kelembapan, tekanan udara, dan kecepatan angin. Algoritma tersebut berhasil diimplementasikan, menghasilkan model dengan 50% data yang digunakan untuk pelatihan dan mencapai akurasi sebesar 60,87%. Namun, dapat disimpulkan bahwa kinerja metode Gaussian Naïve Bayes belum memuaskan untuk klasifikasi cuaca di Kota Tanjungpinang.

REFERENSI

- [1] B. Lakitan, *Dasar-dasar klimatologi*, vol. 175. Raja Grafindo Persada, 1997.
- [2] I. G. A. Gunadi and A. A. K. Dewi, "Klasifikasi curah hujan di provinsi bali berdasarkan metode naïve bayesia," *J. Mat. Sains, dan Pembelajarannya*, pp. 14–25, 2018.
- [3] I. N. Mahmudha, "Klasifikasi Curah Hujan di Kota Semarang Menggunakan Metode Gaussian Naive Bayes," Universitas Sebelas Maret, Surakarta, 2022.
- [4] R. Hutami and E. Z. Astuti, "Implementasi Metode K-Nearest Neighbor Untuk Prediksi Penjualan Furniture Pada Cv. Octo Agung Jepara," *Dokumen Karya Ilmiah*, 2016.
- [5] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Third., vol. 744. Elsevier, 2011.
- [6] F. Handayani and F. S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," *Jurnal Teknik Elektro*, vol. 1, 2015.
- [7] D. T. Larose, *Discovering Knowledge in Data*. Wiley, 2004. doi: 10.1002/0471687545.
- [8] B. Kurniawan, M. A. Fauzi, and A. W. Widodo, "Klasifikasi Berita Twitter Menggunakan Metode Improved Naive Bayes," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 1, pp. 1193–1200, Jul. 2017.
- [9] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," Feb. 2013.
- [10] J. C. Griffis, J. B. Allendorfer, and J. P. Szaflarski, "Voxel-based Gaussian naïve Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans.," *J Neurosci Methods*, vol. 257, pp. 97–108, Jan. 2016, doi: 10.1016/j.jneumeth.2015.09.019.