

Prakiraan Cuaca Kota Bandung di Tahun Berikutnya dengan Menggunakan Metode Naïve Bayes

Afifah Nur Sallamah¹, Adi Sunandar², Alifa Maha Rizka³, Chaerur Rozikin⁴

^{1,2,3,4}Program Studi Informatika, Universitas Singaperbangsa Karawang

2010631170001@student.unsika.ac.id, 20106370041@student.unsika.ac.id,

201063170047@student.unsika.ac.id, chaerur.rozikin@staff.unsika.ac.id

Abstract: Weather is a factor that influences all human activities, making weather prediction crucial for the smooth operation of communities worldwide. This research aims to predict the weather in Bandung city using the Naïve Bayes method. The study utilizes weather data from Bandung city for the past four years (2019-2022), obtained from the BMKG website, including date, minimum temperature, maximum temperature, humidity, wind speed, and rainfall.

Keywords: Weather prediction, Bandung city, Naïve Bayes.

Abstrak: Cuaca menjadi suatu faktor yang mempengaruhi seluruh kegiatan manusia, sehingga prediksi cuaca menjadi sangat diperlukan untuk kelancaran kegiatan masyarakat di dunia. Penelitian ini ditujukan untuk memprediksi cuaca di kota Bandung menggunakan metode *Naïve Bayes*. Penelitian ini menggunakan data cuaca Kota Bandung selama 4 tahun terakhir (2019 - 2022) yang didapatkan dari laman BMKG yang mencakup tanggal, temperatur minimum, temperatur maksimum, kelembapan, kecepatan angin, dan curah hujan.

Kata kunci: Prediksi cuaca, Kota Bandung, *Naïve Bayes*.

Pendahuluan

Meteorologi atau ilmu cuaca adalah ilmu pengetahuan yang membahas peristiwa-peristiwa cuaca dalam jangka waktu dan ruang terbatas. Keadaan cuaca dipengaruhi dari berbagai parameter seperti suhu, tekanan, udara, kecepatan angin, kelembapan udara, dan berbagai fenomena atmosfer lainnya (M. Imam Whidyarto, et al., 2023). Cuaca merupakan suatu keadaan udara disaat tertentu dan wilayah tertentu yang relative sempit serta jangka waktu yang cukup singkat (Fitri Inasni, et al., 2022).

Pada saat ini, faktor cuaca dapat mempengaruhi dalam kehidupan sehari-hari terutama pada bidang transportasi, bidang pertanian, bidang perikanan, bidang perkebunan, bidang pembangunan, dan bidang lainnya. Bagi masyarakat untuk dapat mengetahui setiap informasi terkait cuaca dapat melihat melalui BMKG (Badan Meteorologi Klimatologi dan Geofisika) di setiap provinsi yang ada di Indonesia. Dengan adanya prediksi cuaca, sangat

penting dan dibutuhkan dalam memprakirakan cuaca sebagai langkah antisipasi untuk memperkecil dampak yang akan terjadi. Prediksi tersebut diharapkan memiliki keakuratan tinggi terhadap cuaca agar aktivitas yang dilakukan manusia tidak terganggu misalnya seperti pada pertanian, perkebunan, dan penerbangan yang melakukan kegiatan berdasarkan pada kondisi cuaca, agar kegiatan tersebut dapat berjalan lancar. Prakiraan cuaca dan iklim adalah bagian dalam sistem informasi yang dapat digunakan untuk dapat memantau kondisi alam pada masa yang akan datang. Pada hakekatnya, sistem informasi ini merupakan cara yang dapat dilakukan untuk memaksimalkan dalam memantau, hingga menjadi bentuk evaluasi atau klasifikasi dalam memprediksi cuaca atau iklim (Amril Mutoi Siregar, et al., 2020).

BMKG kota Bandung sendiri, sudah melakukan prakiraan cuaca dengan menggunakan banyak sumber data dan model analisa cuaca, baik data

tersebut dibuat oleh BMKG Pusat atau melalui suatu web yang dikombinasikan dengan beberapa parameter cuaca yang ada di BMKG kota Bandung. Akan tetapi, akurasi dari ketepatan prediksi masih kurang menurut pakar Analisa BMKG. Dalam memprediksi atau prakiraa cuaca pada wilayah di Indonesia dengan hasil yang akurat diperlukan suatu teknologi cuaca, suhu, dan algoritma. Namun pada saat ini, perkembangan yang relative baru dengan tujuan untuk mendapatkan wawasan dan penemuan ilmiah baru dari data pengamatan kita dapat menggunakan sistem *Machine Learning* (ML) pada ilmu alam. *Machine Learning* dapat digunakan dalam meprediksi dan mengolah data yang didapatkan pada masa lalu hingga saat ini dengan menginputkan data yang akan di berikan pada komputer kemudian di proses menggunakan suatu metode misalnya seperti *Naïve Bayes* (M. Imam Whidyarto, et al., 2023).

Metode *Naïve Bayes Classifier* (NBC) merupakan suatu metode untuk mengatasi masalah dengan mencari nilai peluang. Untuk mengklasifikasikannya dapat menggunakan metode statistic dan probabiistik yang telah di usulkan oleh seorang ilmuwan Inggris yaitu Thomas Bayes. Algoritma naïve bayes ini lebih mudah digunakan karena memiliki alur perhitungan yang cukup mudah dengan hanya membutuhkan data pelatihan yang cukup kecil untuk dapat menentukan perkiraan dan parameter yang diperlukan (Azriel Alfian Rizqi, et al., 2022).

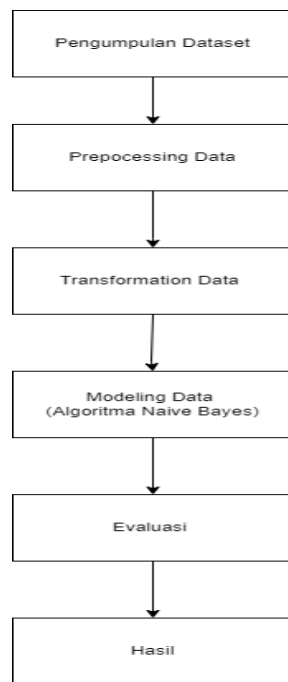
Beberapa penelitian mengenai prakiraan cuaca yang sudah dilakukan sebelumnya yaitu "*Prediksi Cuaca Kota Denpasar Menggunakan Algoritma ELM Dengan Opmtimasi Quantum Delta Particle Swarm Optimization*" yang menunjukkan hasil akurasi dari algortima ELM menggunakan jumlah hidden neuron 5 menghasilkan nilai sebesar 39%, sedangkan dengan optimasi *Quantum Delta Particle Swarm Optimization* dengan parameter, jumlah

partikel 10, jumlah iterasi 42, dan nilai g 0,96 menghasilkan akurasi sebesar 100% (Adam Sulthoni Akbar, et al., 2021). Penelitian selanjutnya "*Prediksi Cuaca Pekanbaru Menggunakna Fuzzy Tsukamoto dan Algoritma Genetika*" menunjukkan nilai akurasi yang didapatkan dari pengujian sistem dengan optimasi *Fuzzy Tsukamoto* dan Algoritma genetika sebesar 72% dengan nilai *probabilitas crossover* 0,6 dan *probabilitas mutation* 0,4. Hasil evaluasi penelitian tersebut membuktikan bahwa kombinasi kedua metode itu mampu menghasilkan hasil akhir yang optimal (Fitri Insani, et al., 2020). Penelitian selanjutnya "*Optimasi Prakiraan Cuaca Menggunakan Metode Ensemble pada Naïve Bayes dan C4.5*" penelitian ini dilakukan untuk meningkatkan akurasi peramalan cuaca dengan model *Naïve Bayes* dan C4.5. Hasil akurasi menggunakan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 49.45% dan algoritma C4.5 menghasilkan akurasi sebesar 41.42% sementara pada proses posttestakurasi yang didapatkan adalah 49,76% untuk bagging naïve Bayes, 46,47% untuk boosting naïve Bayes, 45,76 untuk baggingC4.5 dan 38,82% untuk C4.5 (Vini Indri Yani, et al., 2022).

Berdasarkan penelitian serta penjelasan sebelumnya, penelitian ini bertujuan untuk memprediksi bagaimana cuaca pada tahun berikutnya di Kota Bandung dengan menggunakan algoritma *Naïve Bayes*. Pada penelitian ini, data input dan data uji berasal dari data harian cuaca yang di miliki BMKG (Badan Meteorologi Klimatologi dan Geofisika) Kota Bandung.

Metode Penelitian

Penelitian ini bertujuan untuk memecahkan masalah yang ada dengan menggunakan beberapa metode pada Gambar 1.



Gambar 1. Metode Penelitian

1. Pengumpulan Data: Pada penelitian ini, data yang digunakan berasal dari stasiun cuaca Bandung. Data tersebut mencakup pengamatan cuaca harian selama 4 tahun dan terdiri dari 1.462 data.
2. Preprocessing Data: Sebelum data dapat digunakan untuk pemodelan, dilakukan beberapa langkah preprocessing. Proses ini meliputi pembersihan data dari nilai-nilai yang hilang atau tidak valid, penanganan outlier, dan normalisasi data jika diperlukan.
3. Transformasi Data: Setelah preprocessing, data dapat mengalami transformasi untuk meningkatkan kualitas atau mengungkapkan informasi yang lebih berguna.
4. Pemodelan Data dengan Metode Naive Bayes: Setelah data siap, metode yang digunakan dalam penelitian ini adalah metode Naive Bayes dengan membagi data menjadi data latih dan data uji, pelatihan model Naive Bayes menggunakan data latih, dan Pengujian model menggunakan data uji dan evaluasi performa model.

5. Evaluasi: Setelah model dibangun, dilakukan evaluasi untuk mengukur sejauh mana model dapat memprediksi dengan akurat kelas target pada data yang belum pernah dilihat sebelumnya.
6. Hasil: Setelah evaluasi dilakukan, Hasil dari penelitian ini akan mencakup performa model Naive Bayes dalam memprediksi cuaca berdasarkan data yang digunakan berupa analisis visual yaitu grafik yang menunjukkan perbandingan antara prediksi dan cuaca actual.

Hasil

Setelah mendapatkan data set cuaca kota bandung selama 4 tahun terakhir dari BMKG, kemudian melakukan pemrosesan menggunakan algoritma machine learning. Untuk memprosesnya akan digunakan beberapa library yang diimport yaitu pandas, trait_test_split, GaussianNB, dan matplotlib.pyplot.

```

# import library
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
import matplotlib.pyplot as plt
  
```

Gambar 2. Import Library

Pada gambar 2 menunjukkan beberapa library yang akan digunakan dalam memprediksi cuaca sebagai berikut:

1. Pandas ialah library yang berfungsi sebagai wadah untuk memanipulasi dan menganalisis data.
2. Train_test_split adalah fungsi dari library scikit-learn (sklearn) yang berfungsi untuk membagi dataset menjadi data latih (train set) dan data uji (test set).
3. GaussianNB adalah fungsi yang menggunakan distribusi gaussian untuk menggunakan metode Naive Bayes.
4. Matplotlib.pyplot adalah submodule dari library matplotlib yang digunakan untuk visualisasi grafik dan plot dalam python.

Kemudian setelah library berhasil diimport kita dapat melakukan load pada dataset yang sudah diambil dari laman BMKG untuk cuaca kota bandung selama 4 tahun terakhir (2019-2022) menggunakan fungsi dari pandas dapat kita lihat pada Gambar 3.

```
# load dataset
data = pd.read_csv('clean_weather.csv')
data.head(5)
```

	date	temp_min	temp_max	humidity	wind	rain_rate
0	1/1/2019	22	28	79	2	0
1	2/1/2019	21	31	74	4	14
2	3/1/2019	21	31	76	4	1
3	4/1/2019	20	31	69	3	0
4	5/1/2019	20	32	70	3	0

Gambar 3. Load Dataset

Untuk memastikan data yang akan diolah adalah data yang valid atau tidak ada nilai null maka dapat dilakukan pre-processing data untuk mengecek nilai null pada dataset seperti Gambar 4.

```
# PRE-PROCESSING DATA
# Mengecek nilai null
print(data.isnull())

# Cek kembali data kosong hasil penggabungan
print(data.isnull().sum())
```

	date	temp_min	temp_max	humidity	wind	rain_rate
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...
1456	False	False	False	False	False	False
1457	False	False	False	False	False	False
1458	False	False	False	False	False	False
1459	False	False	False	False	False	False
1460	False	False	False	False	False	False

```
[1461 rows x 6 columns]
date      0
temp_min  0
temp_max  0
humidity  0
wind      0
rain_rate 0
dtype: int64
```

Gambar 4. Preprocessing data

Gambar 4 menunjukkan bahwa data yang digunakan tidak memiliki nilai null pada setia kolomnya. Setelah dipastikan bahwa data tidak ada yang null atau tidak valid, maka selanjutnya data rain_rate diubah nilainya menjadi bentuk string agar memudahkan dalam visualisasi data seperti Gambar 5.

```
# Fungsi kustom untuk mengganti nilai berdasarkan rentang
def replace_range_value(value):
    if value == 0:
        return 'cerah'
    elif 0 < value <= 35:
        return 'hujan ringan'
    elif 35 < value <= 70:
        return 'hujan sedang'
    elif value >= 70:
        return 'hujan deras'
    else:
        return value

# Mengganti nilai integer pada kolom 'rain_rate' dengan nilai string
# berdasarkan rentang menggunakan apply()
data['rain_rate'] = data['rain_rate'].apply(replace_range_value)

# Menampilkan hasil perubahan
print(data)
```

	date	temp_min	temp_max	humidity	wind	rain_rate
0	1/1/2019	22	28	79	2	cerah
1	2/1/2019	21	31	74	4	hujan ringan
2	3/1/2019	21	31	76	4	hujan ringan
3	4/1/2019	20	31	69	3	cerah
4	5/1/2019	20	32	70	3	cerah
...
1456	27-12-2022	19	26	83	2	hujan ringan
1457	28-12-2022	20	26	81	2	cerah
1458	29-12-2022	20	25	84	2	hujan ringan
1459	30-12-2022	20	25	85	2	hujan ringan
1460	31-12-2022	20	25	84	2	hujan ringan

Gambar 5. Mengubah nilai rain_rate

Setelah data sudah berhasil diubah nilainya data dapat dibagi menjadi 2 yaitu sebagai fitur x dan fitur y yang kemudian akan dibagi menjadi 2 lagi untuk data testing dan data training dengan pembagian 75% data training dan 25% data testing pada Gambar 6.

```
# Memisahkan dataset menjadi fitur (X) dan target (y)
X = data[['temp_min', 'temp_max', 'humidity', 'wind']].values
y = data['rain_rate'].values

# Membagi dataset menjadi data training dan data testing
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size=0.25, random_state=0)
```

Gambar 6. Membagi dataset

Kemudian setelah dibagi menjadi 2, data dapat mulai dimasukkan ke dalam metode *naïve bayes* dengan melatihnya menggunakan fitur GaussianNB serta menjalankan testing berdasarkan model GaussianNB yang sudah dilatih menggunakan data training pada Gambar 7.

```
# Melatih model Naive Bayes menggunakan training data
model = GaussianNB()
model.fit(X_train, y_train)

# Memprediksi hasil dari testing data
y_pred = model.predict(X_test)
```

Gambar 7. Melatih model GaussianNB

Setelah data model dilatih dan diuji, kita dapat mengukur tingkat ketepatan akurasi dari model tersebut pada Gambar 8.

```
# Evaluasi akurasi model
accuracy = (y_pred == y_test).mean()
print('Akurasi Model:', accuracy)

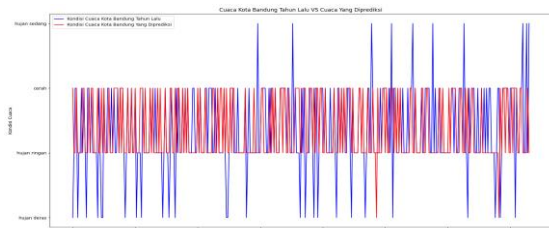
Akurasi Model: 0.6557377049180327
```

Gambar 8. Evaluasi akurasi

Gambar 8 menunjukkan bahwa semakin tinggi nilai akurasi maka semakin akurat data model yang sudah dibuat. Akurasi ini bergantung pada jumlah dataset yang digunakan dan seberapa banyak pembagian pada data training dan data testing. Kemudian, kita dapat memvisualisasikan data untuk model GaussianNB yang sudah dibuat menggunakan fitur dari library matplotlib.pyplot pada Gambar 9.

```
# plot actual vs predicted weather conditions
days = range(len(y_test))
plt.plot(days, y_test, 'b-', label='Kondisi Cuaca Kota Bandung Tahun Lalu')
plt.plot(days, y_pred, 'r-', label='Kondisi Cuaca Kota Bandung Yang Diprediksi')
plt.title('Cuaca Kota Bandung Tahun Lalu VS Cuaca Yang Diprediksi')
plt.xlabel('Hari Ke-')
plt.ylabel('Kondisi Cuaca')
plt.legend()
plt.show()
```

Gambar 9. Membuat visualisasi data



Gambar 9. Visualisasi data

Pembahasan

Setelah penelitian dilakukan dan data berhasil divisualisasikan dengan model dari Gaussian Naïve Bayes, dapat dilihat bahwa cuaca di kota Bandung berada pada rentang cerah dan hujan ringan dan hal ini sejalan dengan tujuan penelitian untuk mencari cara memprediksi cuaca kota Bandung selama 1 tahun ke depan menggunakan data cuaca selama 4 tahun terakhir.

Kesimpulan

Hasil penelitian menunjukkan jika memprakirakan atau memprediksi cuaca menggunakan metode *Naïve Bayes* dengan data yang didapatkan dari BMKG kota Bandung selama 4 tahun terakhir dapat digunakan untuk memprediksikan cuaca selama 1 tahun ke depan dengan ketepatan prediksi dari hasil akurasi sebesar 65%.

Daftar Pustaka

- Azmi, A. U., Hadi, A. F., Dewi, Y. S., Tirta, I. M., Ubaidillah, F., & Anggraeni, D. (2022, February). Naive Bayes Classifier (NBC) for Forecasting Rainfall in Banyuwangi District Using Projection Pursuit Regression (PPR) Method. In *International Conference on Mathematics, Geometry, Statistics, and Computation (IC-MaGeStiC 2021)* (pp. 190-195). Atlantis Press.
- Ahmad Zulfikri R, Gunawan, dan Wresti Andriani (2023, March). Tinjauan Pustaka Sitematis : Penerapan Metode Naive Bayes untuk Klasifikasi dalam Dataset Cuaca. vol.2, no.2, p. 2809-8137.
- Azriel Alfa Rizqi, dan Dewi Kusumaningsih (2022, Spetember). Klasifikasi Curah Hujan di Kota Bogor Provinsi Jawa Barat dengan Menggunakan Metode Naive Bayes.
- Vini Indri Yani, Aradea, dan Husni Mubarak (2022, Desember). Optimasi Prakiraan Cuaca Menggunakan Metode Ensemble pad Naive Bayes dan C4.5. Vol.8, no.3, p : 2443-2210.
- Adam Sulthoni Akbar, Candra Dewi, dan Randy Cahya Wihandika (2021, March). Prediksi cuaca Kota Denpasar menggunakan algoritman ELM dengan Optimasi Quantum Delta Particle Swarm Optimization. vol.5, no.3, 1126-1135.
- Fitri Insani, Syafawani Fadilah, Jasril, dan Suwanto Sanjaya (2020, Desember). Prediksi Cuaca Pekanbaru Menggunakan Fizzy Tsukamoto dan Algoritma Genetika. p: 2597-7271
- Amril Mutoi Siregar, Sutan Faisal, Yana Cahyana, dan Bayu Priyatna. Perbandingan Algoritme Kllasifikasi untuk Predikdi Cuaca.
- Prasetya, R., & Ridwan, A. (2020). Data mining application on weather prediction using classification

- tree, naïve bayes and K-nearest neighbor algorithm with model testing of supervised learning probabilistic brier score, confusion matrix and ROC. *J Appl Commun Inf Technol*, 4(2), 25-33.
- Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather*, 17(8), 1166-1207.
- Kwon, Y., Kwasinski, A., & Kwasinski, A. (2019). Solar irradiance forecast using naïve Bayes classifier based on publicly available weather forecasting variables. *Energies*, 12(8), 1529.
- Yudianto, M. R. A., Agustin, T., James, R. M., Rahma, F. I., Rahim, A., & Utami, E. (2021). Rainfall Forecasting to Recommend Crops Varieties Using Moving Average and Naive Bayes Methods. *International Journal of Modern Education & Computer Science*, 13(3).
- Ali, M. F. M., Askilany, S. A., El-wahab, M. A., & Hassan, M. A. (2019). Data mining algorithms for weather forecast phenomena: comparative study. *International journal of computer science and network security*, 19(9), 76-81.
- Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77, 5198-5219.
- Salmi, N., & Rustam, Z. (2019, June). Naïve Bayes classifier models for predicting the colon cancer. In *IOP conference series: materials science and engineering* (Vol. 546, No. 5, p. 052068). IOP Publishing.
- Devika, R., Avilala, S. V., & Subramaniaswamy, V. (2019, March). Comparative study of classifier for chronic kidney disease prediction using naïve bayes, KNN and random forest. In *2019 3rd International conference on computing methodologies and communication (ICCMC)* (pp. 679-684). IEEE.
- Ali, A., Khairan, A., Tempola, F., & Fuad, A. (2021). Application Of Naïve Bayes to Predict the Potential of Rain in Ternate City. In *E3S Web of Conferences* (Vol. 328, p. 04011). EDP Sciences.
- Rahman, A. U., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., ... & Mosavi, A. (2022). Rainfall prediction system using machine learning fusion for smart cities. *Sensors*, 22(9), 3504.
- Singh, N., Chaturvedi, S., & Akhter, S. (2019, March). Weather forecasting using machine learning algorithm. In *2019 International Conference on Signal Processing and Communication (ICSC)* (pp. 171-174). IEEE.
- Bochenek, B., & Ustrnul, Z. (2022). Machine learning in weather prediction and climate analyses—applications and perspectives. *Atmosphere*, 13(2), 180.
- Cho, D., Yoo, C., Im, J., & Cha, D. H. (2020). Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. *Earth and Space Science*, 7(4), e2019EA000740.
- Verma, G., Mittal, P., & Farheen, S. (2020, March). Real time weather prediction system using IOT and machine learning. In *2020 6th International Conference on Signal Processing and Communication (ICSC)* (pp. 322-324). IEEE.