

Optimasi *Gaussian Naïve Bayes* dengan *Hyperparameter Tuning* dan *Univariate Feature Selection* dalam Prediksi Cuaca

Lindawati¹, Mohammad Fadhli¹, Antony Sandi Wardana^{1,*}

¹ Program Studi Teknik Telekomunikasi, Politeknik Negeri Sriwijaya, Indonesia

* Correspondence: antonisandi400@gmail.com

Copyright: © 2023 by the authors

Received: 2 Agustus 2023 | Revised: 4 Agustus 2023 | Accepted: 6 Agustus 2023 | Published: 20 Desember 2023

Abstrak

Pentingnya melakukan penelitian prediksi cuaca karena perubahan cuaca memiliki pengaruh signifikan dalam kehidupan sehari-hari. Tujuan penelitian ini adalah menerapkan metode klasifikasi *machine learning* yang optimal untuk memprediksi cuaca. Metode yang digunakan adalah model *Gaussian Naïve Bayes* yang telah dioptimasi menggunakan teknik optimasi *Univariate Feature Selection ANOVA-f test* dan *Hyperparameter Tuning GridsearchCV*. Data yang digunakan terdiri dari 6454 data cuaca harian di Kota Palembang. Terdapat 5 uji data model *Gaussian Naïve Bayes* sebelum dan setelah dilakukan optimasi. Hasil penelitian menunjukkan bahwa optimasi model berhasil meningkatkan performa dalam memprediksi cuaca. Hasil akurasi tertinggi setelah optimasi mencapai 98,33% pada 644 data uji, meningkat dari akurasi sebelum optimasi yang hanya 96,95%. Sebelum dilakukan optimasi, hasil prediksi untuk kondisi cuaca cerah berawan, mendung/hujan, hujan sedang, dan hujan lebat sesuai dengan data aktual. Namun, terdapat 20 kesalahan prediksi saat dihadapkan pada data yang seharusnya merupakan kondisi hujan sangat lebat. Setelah dilakukan optimasi, jumlah kesalahan prediksi pada data hujan sangat lebat berhasil berkurang menjadi 7 kesalahan. Pendekatan optimasi yang digunakan dalam penelitian ini membantu mencari kombinasi parameter yang paling tepat, dan fitur-fitur yang tidak relevan dapat dieliminasi, sehingga model hanya mempertimbangkan fitur-fitur yang memiliki kontribusi signifikan dalam memprediksi cuaca.

Kata kunci: prediksi cuaca; *gaussian naïve bayes*; *univariate feature selection*; *hyperparameter tuning*

Abstract

The importance of conducting weather prediction research is due to the significant influence of weather changes on daily life. The purpose of this study is to apply an optimal machine learning classification method for weather prediction. The method used is the *Gaussian Naïve Bayes* model, which has been optimized using *Univariate Feature Selection ANOVA-f test* and *Hyperparameter Tuning GridsearchCV* techniques. The data used consists of 6454 daily weather data in Palembang City. There are 5 tests on the *Gaussian Naïve Bayes* model before and after optimization. The research results show that the optimization of the model successfully improves the performance in weather prediction. The highest accuracy result after optimization reaches 98.33% with 644 test data, an improvement from the pre-optimization accuracy of only 96.95%. Before optimization, the predictions for weather conditions such as sunny, cloudy/rainy, light rain, and heavy rain match the actual data. However, there were 20 prediction errors when dealing with data that should represent very heavy rain conditions. After optimization, the number of prediction errors for the very heavy rain data reduced to seven. The optimization approach used in this research helps find the most suitable parameter combinations and eliminates irrelevant features, allowing the model to consider only significant features in weather prediction.

Keywords: weather prediction; *gaussian naïve bayes*; *univariate feature selection*; *hyperparameter tuning*



PENDAHULUAN

Cuaca adalah keadaan fisik udara dan peristiwa perubahan atmosfer yang terjadi dari waktu ke waktu di suatu tempat dan waktu tertentu. Berbagai faktor, seperti suhu rata-rata, kelembaban, kecepatan angin, arah angin, lama penyinaran matahari, curah hujan, dan lain-lain, dapat mempengaruhi perubahan cuaca (Sunarmi et al., 2022). Informasi cuaca yang akurat memungkinkan masyarakat dan berbagai sektor untuk mengambil langkah yang tepat dan adaptasi yang diperlukan untuk menghadapi perubahan cuaca yang mungkin terjadi, sehingga prediksi cuaca menarik untuk diteliti karena perubahan cuaca memiliki peran vital dalam berbagai sektor, seperti industri, sarana penerbangan, pertanian serta kegiatan sehari-hari seperti perencanaan perjalanan, olahraga dan sebagainya (Intan & Koswara, 2021).

Salah satu cara yang dapat digunakan untuk melakukan prediksi cuaca adalah dengan menggunakan kecerdasan buatan. Tujuan utama kecerdasan buatan adalah untuk memahami dan memodelkan proses berpikir manusia, serta untuk membuat algoritma yang dapat digunakan untuk memodelkan proses berpikir manusia yang menjadi salah satu bagian dari *artificial intelligence* yaitu *machine learning* (Harianto et al., 2021). Namun terdapat tantangan dalam meningkatkan akurasi prediksi cuaca menggunakan *machine learning*. Beberapa masalah yang dihadapi adalah data yang sedikit, kompleksitas dan variasi data cuaca harian, termasuk suhu, kelembaban udara, kecepatan angin, dan kondisi cuaca yang beragam (Utami et al., 2021). Selain itu, ada kemungkinan adanya fitur-fitur cuaca yang tidak relevan atau memiliki tingkat korelasi yang rendah dengan kelas cuaca yang diinginkan, sehingga dapat mempengaruhi akurasi prediksi (Yani et al., 2022).

Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan dua pendekatan metode dalam *machine learning* yaitu *Hyperparameter Tuning GridsearchCV* dan *Univariate Feature Selection ANOVA-f test* dalam optimasi model *Gaussian Naive Bayes*. *Hyperparameter tuning* akan membantu mencari kombinasi parameter yang paling tepat dalam model *Gaussian Naive Bayes* untuk meningkatkan kemampuan dalam memprediksi (Ardiansyah et al., 2021; Armansyah & Ramli, 2022) cuaca dengan lebih akurat. Sementara itu, *Feature Selection ANOVA-f test* akan membantu mengidentifikasi fitur-fitur cuaca yang memiliki korelasi tinggi dengan kelas cuaca yang diinginkan, sehingga dapat mengurangi dimensi fitur yang tidak relevan dan meningkatkan efisiensi prediksi (Harpale & Bairagi, 2021; Jebadurai et al., 2022). Dengan melakukan pendekatan tersebut, penelitian yang mereka lakukan memberikan kontribusi dalam pengembangan teknik prediksi cuaca yang lebih efisien dan efektif dalam mengatasi kompleksitas dan variasi data cuaca harian, sehingga informasi cuaca yang lebih handal dapat digunakan untuk perencanaan dan pengambilan keputusan yang lebih baik.

Hasil temuan sebelumnya menunjukkan bahwa beberapa algoritma *machine learning* telah digunakan untuk prediksi cuaca, namun performanya masih belum optimal. Beberapa algoritma mencapai akurasi yang rendah, seperti *Naive Bayes* dan *C4.5* dengan metode *Ensemble Learning* yang hanya mencapai akurasi sekitar 49,45% dan 41,24% (Yani et al., 2022). Meskipun terdapat algoritma dengan akurasi lebih tinggi, seperti *Deep Learning* dengan akurasi 82,92%, *Generalized Linear Model* dengan akurasi 84,06% (Siregar, 2020). Dan penerapan algoritma *K-Nearest Neighbor* (KNN) dengan akurasi 89% (Rangkuti et al., 2021), masih ada ruang untuk peningkatan akurasi prediksi cuaca. Semakin tinggi tingkat akurasi dari suatu algoritme, semakin baik kualitas prediksinya. Akurasi menggambarkan seberapa dekat hasil prediksi dengan nilai sebenarnya. Jika akurasi semakin tinggi, artinya algoritma dapat melakukan prediksi dengan tingkat ketepatan yang lebih tinggi, dan hasil prediksinya lebih mendekati kenyataan (Siregar, 2020).

Selain masalah akurasi, penelitian sebelumnya juga belum sepenuhnya mencakup analisis tentang efisiensi dan relevansi fitur dalam prediksi cuaca. Dataset yang kurang, fitur-fitur cuaca yang tidak relevan atau memiliki tingkat korelasi rendah dengan kelas cuaca yang

diinginkan dapat mempengaruhi akurasi prediksi (Yani et al., 2022; Utami et al., 2021). Oleh karena itu, diperlukan pendekatan yang dapat mengidentifikasi fitur-fitur penting dan relevan sehingga dapat mengurangi dimensi fitur yang tidak relevan dan meningkatkan efisiensi prediksi seperti yang disarankan pada penelitian (Oshodi, 2022).

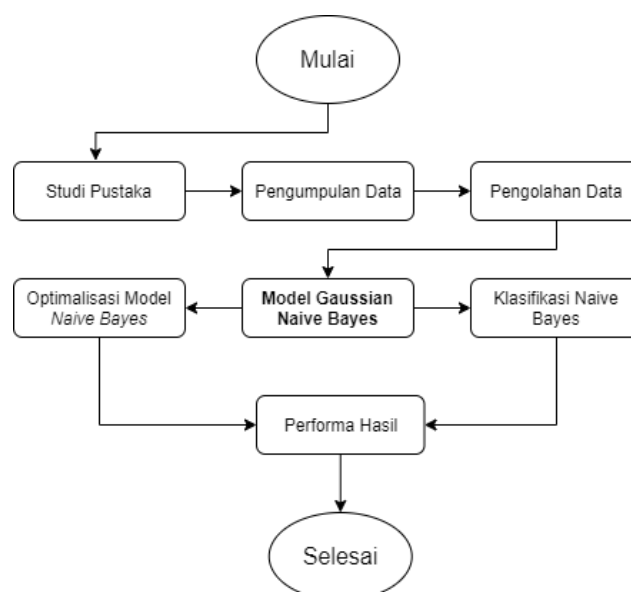
Hasil temuan sebelumnya juga menunjukkan bahwa prediksi cuaca menggunakan *Gaussian Naive Bayes* banyak digunakan (Siregar et al., 2020 ; Yani et al., 2022), dan memiliki kinerja terbaik dengan prediksi akurasi sebesar 84,153% pada penelitian (Oshodi, 2022). Namun, metode *Hyperparameter Tuning* dan *Feature Selection ANOVA-f test* dalam model *Gaussian Naive Bayes* belum banyak digunakan dalam penelitian sebelumnya, dengan melakukan pendekatan ini, diharapkan dapat mengatasi kelemahan dan kekurangan dari penelitian sebelumnya dan menghasilkan prediksi cuaca yang lebih akurat dan efisien.

Penelitian yang dilakukan ini bertujuan untuk mengatasi kompleksitas dan variasi data cuaca yang beragam dengan menggunakan dua pendekatan metode dalam *machine learning*, yaitu *Hyperparameter Tuning* dan *Feature Selection ANOVA-f test* dalam model *Gaussian Naive Bayes*. Selain itu, dengan diterapkan metode ini untuk meningkatkan akurasi prediksi cuaca sehingga informasi cuaca yang lebih handal dapat digunakan untuk perencanaan dan pengambilan keputusan yang lebih baik dalam berbagai sektor dan aktivitas sehari-hari, serta diharapkan hasil prediksi cuaca menjadi lebih akurat, sehingga pada penelitian yang dilakukan ini akan menunjukkan seberapa besar pengaruh optimasi dengan metode yang diterapkan di atas terhadap model *Gaussian Naive Bayes* menggunakan dataset cuaca harian Kota Palembang dalam meningkatkan performa akurasi prediksi cuaca.

METODE

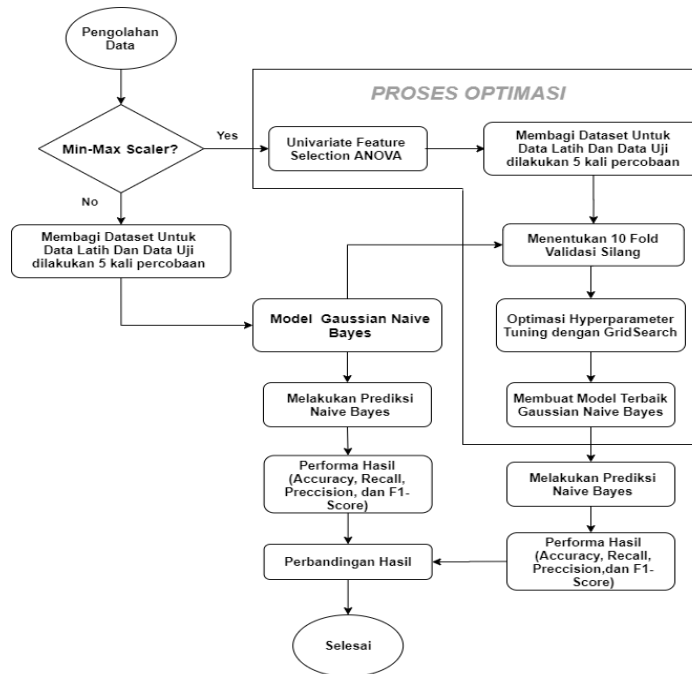
Penelitian ini, dilakukan pengujian untuk meningkatkan akurasi prediksi cuaca di Kota Palembang menggunakan model *Gaussian Naive Bayes*. *Gaussian Naive Bayes* akan diimplementasikan menggunakan bahasa pemrograman *Python*. Pemilihan bahasa pemrograman ini penting karena akan memastikan program yang dikembangkan berfungsi secara optimal (Sari et al., 2020).

Model *Gaussian Naive Bayes* akan dioptimasi dengan menggunakan metode *Univariate Features Selection ANOVA-f test* dan *Hyperparameter GridsearchCV*. Selanjutnya, dilakukan perbandingan performa model sebelum dan setelah dilakukan optimasi untuk melihat seberapa besar peningkatan akurasi yang berhasil dicapai. Alur penelitian bisa dilihat pada gambar 1.



Gambar 1. Alur penelitian

Studi pustaka adalah proses mempelajari dan mengumpulkan informasi dari berbagai sumber tentang topik penelitian untuk mendapatkan pemahaman yang komprehensif, yang membantu dalam perencanaan dan pelaksanaan penelitian (Adlini et al., 2022). Pengumpulan data untuk dijadikan dataset penelitian menggunakan data cuaca harian kota Palembang dalam rentang waktu Januari 2000 - April 2023 berjumlah 8400 data pada Stasiun Cuaca klimatologi Palembang yang bersumber dari website <https://dataonline.bmkg.go.id>.



Gambar 2. Proses pengujian model *gaussian naïve bayes*

Pada gambar 2 Pengolahan data (*Pre-processing*) dilakukan dengan memilih variabel independen dan dependen, menghapus data yang tidak diperlukan, membuat label pada data kategorikal, dan menghapus data yang kosong (Hendrawan et al., 2022; Sihombing et al., 2021; Sudrajat et al., 2022). Setelah dilakukan pre-processing data, didapatkan dataset yang terdiri dari 9 variabel independen (Suhu Min, Suhu Max, Suhu Rata, Kel_Rata, LPM, Kec_Angin Max, Arah Angin, Kec_Angin Rata, dan Curah Hujan dan 1 variabel dependen (Kondisi Cuaca). Kemudian dataset tersebut dijadikan sebagai data latih dan data uji dengan rasio 90:10, 80:20, 70:30, 60:40, 50:50. Data latih digunakan untuk mendapatkan nilai pada tabel probabilitas dan data uji digunakan untuk menguji tabel probabilitas yang dihasilkan oleh data latih (Susanti et al., 2022). Langkah berikutnya adalah membuat model *Gaussian Naïve Bayes* dan mengoptimasi menggunakan metode *Univariate Features Selection ANOVA-f test*. Metode ini memilih fitur terbaik berdasarkan *test statistic univariate*, di mana *ANOVA-f test* digunakan untuk menghitung skor fitur berdasarkan jarak antara kelas dan dalam kelas. Kemudian, *ANOVA-f test* menghitung hubungan linier antara fitur dan target, dan fitur dengan korelasi tertinggi dipilih (Jebadurai et al., 2022). Setelah pemilihan fitur, dilakukan optimasi menggunakan *Hyperparameter Tuning GridsearchCV* untuk mencari model terbaik. Proses *Cross-Validation* dengan *10 fold* digunakan, di mana model dilatih dan dievaluasi pada kombinasi subset yang berbeda-beda. Keuntungan dari metode ini adalah setiap subset menjadi data uji tepat satu kali dan sebagai data pelatihan sebanyak 9 kali (Aguni et al., 2021; Alhakeem et al., 2022). Setelah peforma hasil didapatkan, hasil dari sebelum dan sesudah dioptimasi akan di bandingkan berdasarkan tingkat *Accuracy, Precision, Recall, dan F1-Score*.

HASIL DAN PEMBAHASAN

Hasil

Pada penelitian ini, dilakukan optimasi algoritma *Naïve Bayes* menggunakan metode *Univariate Feature Selection ANOVA-f test* dan *Hyperparameter Tuning GridSearchCV* untuk meningkatkan akurasi prediksi cuaca di Kota Palembang. Penggunaan data set cuaca harian kota Palembang awalnya berjumlah 8400, namun setelah dilakukan pembersihan data, seperti menghapus data kosong (*NaN*), menghapus data duplikat, dan menghapus data yang tidak diperlukan pada data kategorikal dalam fitur (Tanggal dan AAT), data set berkurang menjadi 6545 data dengan 9 fitur dan 1 target untuk uji percobaan optimasi model *Naïve Bayes*, yang ditunjukkan pada gambar 3.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8399 entries, 0 to 8398
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Tanggal      8399 non-null   object
1   Suhu Min     8258 non-null   float64
2   Suhu Max     8299 non-null   float64
3   Suhu Rata    8310 non-null   float64
4   Kel_Rata     8271 non-null   float64
5   LPM          8255 non-null   float64
6   Kec_Angin Max 8351 non-null   float64
7   Arah Angin   8285 non-null   float64
8   Kec_Angin Rata 8353 non-null   float64
9   AAT          8330 non-null   object
10  Curah Hujan  6828 non-null   float64
11  Kondisi Cuaca 8399 non-null   object
dtypes: float64(9), object(3)
memory usage: 787.5+ KB

<class 'pandas.core.frame.DataFrame'>
Int64Index: 6545 entries, 0 to 8396
Data columns (total 10 columns):
#   Column      Non-Null count  Dtype
---  ---
0   Suhu Min     6545 non-null   float64
1   Suhu Max     6545 non-null   float64
2   Suhu Rata    6545 non-null   float64
3   Kel_Rata     6545 non-null   float64
4   LPM          6545 non-null   float64
5   Kec_Angin Max 6545 non-null   float64
6   Arah Angin   6545 non-null   float64
7   Kec_Angin Rata 6545 non-null   float64
8   Curah Hujan  6545 non-null   float64
9   Kondisi Cuaca 6545 non-null   object
dtypes: float64(9), object(1)
memory usage: 562.5+ KB
None
    
```

Gambar 3. Data mentah dan data setelah dilakukan pembersihan untuk pengolahan data

Pada proses pemilihan fitur menggunakan metode *Univariate Feature Selection ANOVA-f test* dari 9 fitur yang dipilih, terdapat 3 fitur dengan bobot rendah, yaitu *Kel_Rata*, *Kec_Angin Max*, dan *Arah Angin* seperti yang ditunjukkan pada gambar 4, lalu dipilih 6 fitur terbaik yang memiliki relevansi tinggi dengan target prediksi kemungkinan kondisi cuaca (*Cerah Berawan*, *Mendung* atau *Hujan Ringan*, *Hujan Sedang*, *Hujan Lebat*, dan *Hujan Sangat Lebat*) sehingga dapat mempengaruhi performa model *Gaussian Naïve Bayes*. Selanjutnya, dilakukan *Hyperparameter Tuning* untuk mengetahui parameter yang optimal untuk model *Gaussian Naïve Bayes*. Proses *Cross-Validation* menggunakan *GridSearchCV* digunakan untuk mengoptimalkan akurasi (Maisat & Ashafidz F.D, 2023). Hasil performa dari percobaan tersebut terlihat pada tabel 1 dan tabel 2, di mana data set dibagi menjadi data uji dan data latih sebanyak 5 kali.

	Feature	Score
0	Suhu Min	175.415480
1	Suhu Max	172.658541
2	Suhu Rata	277.677236
3	Kel_Rata	14.411820
4	LPM	120.346672
5	Kec_Angin Max	14.152377
6	Arah Angin	50.707172
7	Kec_Angin Rata	65.305416
8	Curah Hujan	15219.315846

Fitur Terbaik yang dipilih sebagai optimasi:
 Index(['Suhu Min', 'Suhu Max', 'Suhu Rata', 'LPM', 'Kec_Angin Rata',
 'Curah Hujan'],
 dtype='object')

Gambar 4. Hasil pemilihan fitur terbaik

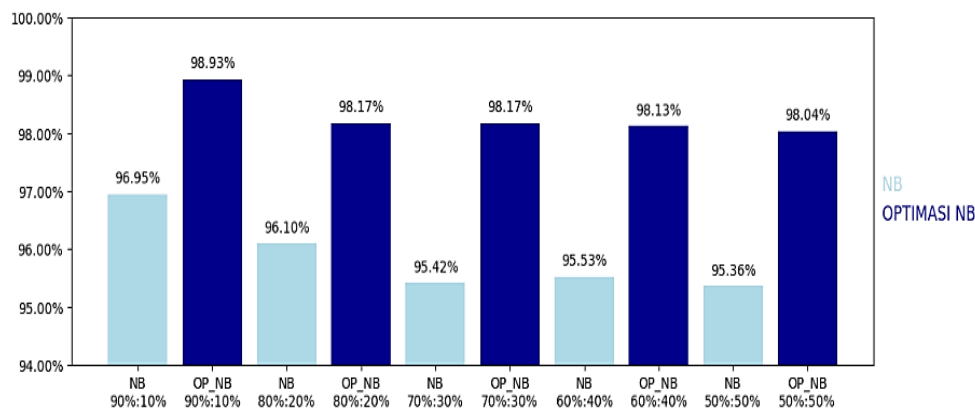
Tabel 1. Rekapitulasi hasil percobaan model naïve bayes sebelum dioptimasi

Percobaan	Accuracy	Precision	Recall	F1-Score
90%:10%	96,95%	95,56%	98,63%	96,79%
80%:20%	96,10%	94,05%	98,13%	95,63%
70%:30%	95,42%	87,98%	91,96%	89,37%
60%:40%	95,53%	87,67%	92,89%	89,79%
50%:50%	95,36%	86,55%	92,53%	88,85%

Tabel 2. Rekapitulasi hasil percobaan model naïve bayes sesudah di optimasi

Percobaan	Accuracy	Precision	Recall	F1-Score
90%:10%	98,93%	98,18%	99,52%	98,80%
80%:20%	98,70%	97,38%	99,30%	98,28%
70%:30%	98,17%	93,32%	93,50%	93,25%
60%:40%	98,13%	92,21%	94,36%	93,21%
50%:50%	98,04%	92,10%	93,96%	92,97%

Hasil rekapitulasi nilai pada tabel 1 dan tabel 2 menunjukkan percobaan menggunakan model *Gaussian Naïve Bayes* sebelum dioptimasi, dan sesudah dioptimasi yang dilakukan sebanyak 5 kali pada data set. Dalam percobaan tersebut, nilai akurasi tertinggi diperoleh ketika data uji digunakan sebesar 10% dan data latih sebesar 90%, dengan akurasi mencapai 96,95% sebelum dioptimasi. dan mendapatkan peningkatan menjadi 98,93% setelah dioptimasi. Selain itu, nilai *Precision*, *Recall*, dan *F1-Score* tertinggi juga didapatkan pada kondisi yang sama yaitu dengan data uji sebesar 10% dan data latih sebesar 90%, nilai-nilai tersebut tertera pada tabel 1, dan tabel 2. Langkah selanjutnya adalah melakukan perbandingan nilai akurasi dari semua percobaan. Hasil perbandingan menunjukkan bahwa nilai akurasi tertinggi, yaitu 98,93%, diperoleh pada model *Gaussian Naïve Bayes* dioptimasi dengan *Univariate Feature Selection ANOVA-f test* dan *Hyperparameter Tuning GridsearchCV* pada rasio perbandingan dengan 10% data uji dan 90% data latih. Hasil perbandingan nilai akurasi dapat dilihat pada gambar 5.



Gambar 5. Perbandingan seluruh hasil performa *accuracy* sebelum dioptimasi dan sesudah dioptimasi

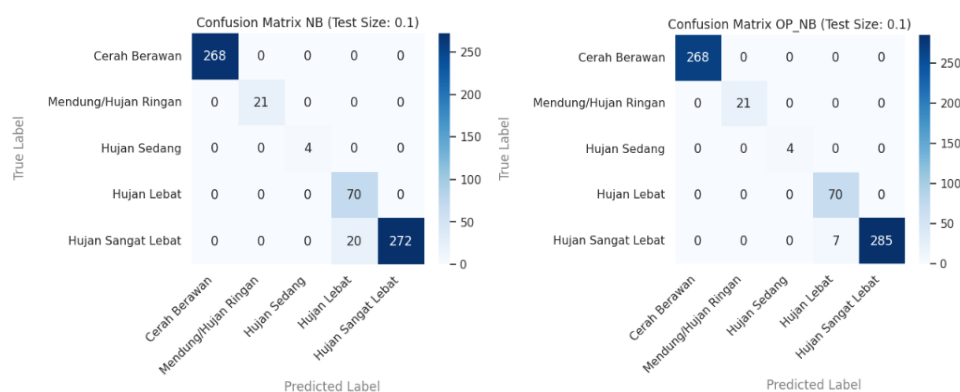
Perbandingan selisih dan peningkatan nilai akurasi yang didapatkan menggunakan model *Gaussian Naïve Bayes* sebelum dioptimasi dan model *Gaussian Naïve Bayes* dioptimasi dengan *Univariate Feature ANOVA* dan *Hyperparameter Tuning GridSearchCV* ditunjukkan pada tabel 3. Selisih peningkatan akurasi tertinggi terjadi pada percobaan menggunakan data latih 70% dan data uji 30% dengan selisih peningkatan sebesar 2,75%.

Tabel 3. Rekapitulasi hasil akurasi seluruh percobaan

Percobaan	NB	OP_NB	Selisih Peningkatan
90%:10%	96,95%	98,93%	1,98%
80%:20%	96,10%	98,70%	2,6%
70%:30%	95,42%	98,17%	2,75%
60%:40%	95,53%	98,3%	2,6%
50%:50%	95,36%	98,04%	2,68%

Hasil pengujian optimasi yang dilakukan, akurasi tertinggi yang didapatkan pada 10% data uji dan 90% data latih dengan akurasi setelah dioptimasi sebesar 98,93%, dari sebelumnya sebelum dioptimasi sebesar 96,95%, terlihat pada gambar 4 hasil prediksi yang didapatkan sehingga mendapatkan akurasi tertinggi dari 10% data uji yang berjumlah 655 data yang diprediksi dengan jumlah data pada kondisi cuaca cerah berawan sebanyak 268 data, pada mendung/hujan ringan sebanyak 21 data, hujan sedang sebanyak 4 data, hujan lebat sebanyak 70 data, dan hujan sangat lebat sebanyak 292 data. Kolom true label adalah data aktual yang terjadi dari stasiun dan predict label adalah hasil prediksi yang dilakukan pada penelitian ini.

Hasil prediksi yang didapatkan penelitian saat menggunakan naïve bayes sebelum dioptimasi pada kondisi cerah berawan, mendung/hujan, hujan sedang dan hujan lebat hasil prediksi benar dengan data actual yang terjadi, namun saat terdapat data yang seharusnya kondisi hujan sangat lebat hasil prediksi sebelum dioptimasi terjadi 20 kesalahan prediksi, setelah dilakukan optimasi dengan menggunakan metode *Univariate Feature Selection* ANOVA-f test dan Hyperparameter Tuning GridsearchCV, hasil 20 kesalahan prediksi pada data hujan sangat lebat berkurang menjadi 7 kesalahan saat setelah dioptimasi.

**Gambar 6.** Confusion matrix hasil prediksi gaussian naïve bayes sebelum di optimasi

Pembahasan

Pengujian yang dilakukan pada penelitian ini dilakukan sebanyak 10 kali percobaan pada masing-masing data set. di pecah menjadi 5 kali percobaan sebelum dioptimasi dan 5 kali setelah dioptimasi dengan membagi data set menjadi data uji dan data latih 10%:90%, 20%:80%, 30%:70%, 40%:60% dan 50%:50%. Percobaan data set tanpa optimasi mendapatkan akurasi tertinggi 96,95% dengan menggunakan 90% data latih dan 10% data uji, setelah dioptimasi akurasi prediksi meningkat 1.98% menjadi 98,93% dengan data latih dan data uji yang sama 90%:10% pada percobaan optimasi ini menjadi akurasi tertinggi dari 10 percobaan yang dilakukan. Pada tabel 3 selisih peningkatan juga terjadi pada percobaan lainnya yaitu 20%:80%, 30%:70%, 40%:60% dan 50%:50%, dengan menggunakan metode *Univariate Feature Selection* ANOVA-f test dan *Hyperparameter Tuning GridsearchCV*, dengan hasil prediksi seperti pada hasil confusion matrix gambar 4 dan gambar 5. Perbandingan seluruh

nilai akurasi bisa dilihat pada gambar 6, tidak hanya peforma *Accuracy* peningkatan juga terjadi dalam peforma *Precision*, *Recall*, dan *F1-Score* pada model *Gaussian Naïve Bayes* sebelum dioptimasi dan sesudah dioptimasi ditunjukkan pada tabel 1 dan tabel 2.

Pada penelitian ini, hasil prediksi menggunakan *Naïve Bayes* sebelum dioptimasi untuk kondisi cuaca cerah berawan, mendung/hujan, hujan sedang, dan hujan lebat menunjukkan prediksi yang benar sesuai dengan data sebenarnya. Namun, saat ada data yang seharusnya merupakan kondisi hujan sangat lebat, terdapat 20 kesalahan prediksi sebelum dilakukan optimasi. Setelah melakukan optimasi dengan menggunakan metode *Univariate Feature Selection ANOVA-f test* dan *Hyperparameter Tuning GridsearchCV*, jumlah kesalahan prediksi pada data hujan sangat lebat berkurang menjadi 7 kesalahan. Perubahan ini dapat dilihat dalam gambar 4. Dengan demikian, hasil optimasi telah berhasil meningkatkan performa prediksi untuk kondisi hujan sangat lebat dan mengurangi jumlah kesalahan prediksi yang sebelumnya terjadi.

Tabel 4. Rekapitulasi hasil prediksi dari 655 data prediksi

<i>Kondisi Cuaca</i>	<i>Hasil Prediksi</i>	<i>Data Aktual</i>	<i>Kesalahan Prediksi</i>
Cerah Berawan NB	268	268	0
Cerah Berawan OP_NB	268	268	0
Hujan Ringan NB	21	21	0
Hujan Ringan OP_NB	21	21	0
Hujan Sedang NB	4	4	0
Hujan Sedang OP_NB	4	4	0
Hujan Lebat NB	70	70	0
Hujan Lebat OP_NB	70	70	0
Hujan Sangat Lebat NB	272	292	20
Hujan Sangat Lebat OP_NB	285	292	7

Hasil temuan kami, *Gaussian Naive Bayes* memiliki kelebihan yang mampu menangani tugas klasifikasi dengan lebih dari dua kelas dan dapat diperluas dengan mudah untuk mengatasi masalah multikelas. (Widyassari & Suryani, 2021; Muhamad et al., 2017) . Meskipun memiliki beberapa kelebihan, *Gaussian Naive Bayes* juga memiliki beberapa kelemahan, seperti kesulitan dalam menangani ketergantungan antar fitur dan ketidakmampuannya memodelkan hubungan yang rumit antara fitur (Harianto et al., 2020). Metode *Hyperparameter tuning GridsearchCV* dan *Univariate Feature Selection ANOVA-f test* dalam model *Gaussian Naive Bayes* dapat menjadi solusi dan memberikan hasil yang baik dibandingkan dengan temuan sebelumnya (Yani et al., 2022), karena penelitian tersebut memiliki kendala yaitu banyak *variance* yang nilainya hampir sama untuk setiap kategori, sehingga sulit untuk dimodelkan bahkan setelah dilakukan optimasi oleh metode *ensemble* yang dilakukan hasil temuan penelitian sebelumnya menunjukkan akurasi sebesar 49,76% dengan dataset sebanyak 3652 data dalam rentang waktu 10 tahun. Namun, penelitian yang dilakukan ini menggunakan dataset yang lebih besar, yaitu sebanyak 6454 data dalam rentang waktu 20,5 tahun, dataset yang lebih banyak karena saran yang diberikan pada penelitian (Utami et al., 2021) yang memiliki hasil akurasi 70% menggunakan metode LSH dan KNN.

Hasil penelitian yang dilakukan ini menunjukkan bahwa akurasi prediksi cuaca meningkat menjadi 98,93%, lebih tinggi dari penelitian sebelumnya. Dengan membandingkan hasil ini dengan temuan sebelumnya, penelitian yang dilakukan ini menunjukkan adanya peningkatan akurasi yang signifikan dan perluasan cakupan dataset, yang mengindikasikan kebaruan dan kontribusi penelitian dalam bidang prediksi cuaca menggunakan model

Gaussian Naïve Bayes dioptimasi dengan metode *Hyperparameter Tuning GridsearchCV* dan *Univariate Feature Selection ANOVA-f test*, sehingga penelitian yang dilakukan ini berhasil meningkatkan performa model dan memberikan pemahaman yang lebih luas tentang prediksi cuaca dengan pendekatan *machine learning*.

SIMPULAN

Penggunaan teknik optimasi *Hyperparameter Tuning GridsearchCV* dan *Univariate Feature Selection ANOVA-f test* terbukti efektif dalam meningkatkan akurasi prediksi cuaca pada model *Gaussian Naive Bayes*. Temuan ini menunjukkan hasil prediksi saat menggunakan *Naïve Bayes* sebelum dioptimasi pada kondisi cerah berawan, mendung/hujan, hujan sedang dan hujan lebat mendapat hasil prediksi benar dengan data aktual yang terjadi. Namun saat terdapat terjadi 20 kesalahan prediksi pada kondisi cuaca hujan sangat lebat, setelah dilakukan optimasi dengan menggunakan metode *Univariate Feature Selection ANOVA-f test* dan *Hyperparameter Tuning GridsearchCV*, hasil 20 kesalahan prediksi pada data hujan sangat lebat berkurang menjadi 7 kesalahan saat setelah dioptimasi. Penggunaan *Hyperparameter Tuning* membantu mencari kombinasi parameter yang optimal untuk model *Gaussian Naive Bayes*. Selain itu, fitur-fitur yang tidak relevan dapat dieliminasi, sehingga model hanya mempertimbangkan fitur-fitur yang berkontribusi signifikan dalam melakukan prediksi cuaca.

REFERENSI

- Adlini, Anisya H.D, Sarah Y, Octavia C, & Sauda J.M. (2022). Metode Penelitian Kualitatif Studi Pustaka. *Edumaspul: Jurnal Pendidikan*, 6(1), 974–980. <https://doi.org/10.33487/edumaspul.v6i1.3394>
- Aguni, L., Chabaa, S., Ibnyaich, S., & Zeroual, A. (2021). Predicting the notch band frequency of an ultra-wideband antenna using artificial neural networks. *Telkonnika (Telecommunication Computing Electronics and Control)*, 19(1), 1–8. <https://doi.org/10.12928/telkonnika.v19i1.15912>
- Alhakeem, Z. M., Jebur, Y. M., Henedy, S. N., Imran, H., Bernardo, L. F., & Hussein, H. M. (2022). Prediction of ecofriendly concrete compressive strength using gradient boosting regression tree combined with GridSearchCV hyperparameter-optimization techniques. *Materials*, 15(21), 7432. <https://doi.org/10.3390/ma15217432>
- Ardiansyah, M., Sunyoto, A., & Luthfi, E. T. (2021). Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4. 5 untuk Klasifikasi Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 147-156. <https://doi.org/10.29408/edumatic.v5i2.3424>
- Armansyah, A., & Ramli, R. K. (2022). Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 1-10. <https://doi.org/10.29408/edumatic.v6i1.4789>
- Hariato, H., Sunyoto, A., & Sudarmawan, S. (2020). Optimasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Anomaly dengan Univariate Fitur Selection. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 40–49. <https://doi.org/10.29408/edumatic.v4i2.2433>
- Harpale, V., & Bairagi, V. (2021). An adaptive method for feature selection and extraction for classification of epileptic EEG signal in significant states. *Journal of King Saud University - Computer and Information Sciences*, 33(6), 668–676. <https://doi.org/10.1016/j.jksuci.2018.04.014>
- Hendrawan, I. R., Utami, E., & Hartanto, A. D. (2022). Comparison of Naïve Bayes Algorithm and XGBoost on Local Product Review Text Classification. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 143-149. <https://doi.org/10.29408/edumatic.v6i1.5613>
- Intan, I., & Koswara, A. T. (2021). Analisis performansi prakiraan cuaca menggunakan algoritma machine learning. *Jurnal_Pekommas*, 6(2), 1-8.

- Jebadurai, I. J., Paulraj, G. J. L., Jebadurai, J., & Silas, S. (2022). Experimental Analysis of Filtering-Based Feature Selection Techniques for Fetal Health Classification. *Serbian Journal of Electrical Engineering*, 19(2), 207–224. <https://doi.org/10.2298/SJEE2202207J>
- Maisat, & Ashafidz F.D. (2023). Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM. *Online) Teknologi: Jurnal Ilmiah Sistem Informasi*, 13(1), 8–15.
- Muhamad, H., Prasojo, C. A., Sugianto, N. A., Surtiningsih, L., & Cholissodin, I. (2017). Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris. 4(3), 180–184. <https://doi.org/10.25126/jtiik.201743251>
- Oshodi. (2022). Machine Learning-based Algorithms for Weather Forecasting. *Preprints*, 1(1), 1–6. <https://doi.org/10.20944/preprints202206.0428.v>
- Rangkuti, M. Y. R., Alfansyuri, M. V., & Gunawan, W. (2021). Penerapan Algoritma K-Nearest Neighbor (Knn) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia. *Hexagon*, 2(2), 11-16. <https://doi.org/10.36761/hexagon.v2i2.1082>
- Sari, V., Firdausi, F., & Azhar, Y. (2020). Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 1–9. <https://doi.org/10.29408/edumatic.v4i2.2202>
- Sihombing, L. O., Hannie, H., & Dermawan, B. A. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 233-242. <https://doi.org/10.29408/edumatic.v5i2.4089>
- Siregar, A. M., Faisal, S., Cahyana, Y., & Priyatna, B. (2020). Perbandingan Algoritme Klasifikasi Untuk Prediksi Cuaca. *Jurnal Accounting Information System (AIMS)*, 3(1), 15-24. <https://doi.org/10.32627/aims.v3i1.280>
- Siregar. (2020). Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning. *PETIR*, 13(2), 138–147. <https://doi.org/10.33322/petir.v13i2.998>
- Sudrajat, A., Mulyani, N., & Marpaung, N. (2022). Sistem Pendukung Keputusan Penentuan Kelayakan Penangguhan Kredit Nasabah menggunakan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 6(2), 205-214. <https://doi.org/10.29408/edumatic.v6i2.6298>
- Sunarmi, N., Kumailia, E. N., Nurfaiza, N., Nikmah, A. K., Aisyah, H. N., Sriwahyuni, I., & Lailly, S. N. (2022). Analisis Faktor Unsur Cuaca terhadap Perubahan Iklim Di Kabupaten Pasuruan pada Tahun 2021 dengan Metode Principal Component Analysis. *Newton-Maxwell Journal of Physics*, 3(2), 56–64. <https://doi.org/10.33369/nmj.v3i2.23380>
- Susanti, S., Sari, A. A., Anam, M. K., Jamaris, M., & Hamdani, H. (2022). Sistem Prediksi Keuntungan Influencer Pengguna E-Commerce Shopee Affiliates menggunakan Metode Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 6(2), 394–403. <https://doi.org/10.29408/edumatic.v6i2.6787>
- Utami, Rini, D. P., & Lestari, E. (2021). Prediksi Cuaca di Kota Palembang Berbasis Supervised Learning Menggunakan Algoritma K-Nearest Neighbour. *JUPITER: Jurnal Penelitian Ilmu Dan Teknologi Komputer*, 13(1), 09–18.
- Widyassari, A. P., & Suryani, P. E. (2021). Komparasi Metode Naïve Bayes dan SAW untuk Pemilihan Penerimaan Insentif Karyawan. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 3(02), 149–159. <https://doi.org/10.46772/intech.v3i02.555>
- Yani, Aradea, & Husni. (2022). Optimizing Weather Forecast Using Ensemble Method on Naïve Bayes and C4.5. *JuTISI*, 8(3), 607–619. <https://doi.org/10.28932/jutisi.v8i3.5455>