



Klasifikasi dan Prediksi Kelembapan Udara Optimal

Perkembangbiakan Nyamuk *Aedes aegypti*

Elysia Putri Linda Triana¹, Putri Febiyanti², Azwar Muhtar³, Robert Kurniawan⁴

^{1,2,3}Program Studi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

⁴Program Studi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia

¹212112017@stis.ac.id

³212111942@stis.ac.id

⁴robertk@stis.ac.id

Corresponding author email: 212112292@stis.ac.id

Abstract: One of the diseases that is still a concern in Indonesia is Dengue Haemorrhagic Fever (DHF). Bandung City is the area with the most DHF cases in Indonesia in the period January-April 2024. DHF is spread by the *Aedes aegypti* mosquito. The mosquito breeds optimally at a humidity of 70–90%. This study aims to classify the optimal air humidity for the breeding of *Aedes aegypti* mosquitoes. The data used comes from BMKG. There are three methods used in the classification, namely Naïve Bayes, SVM, and Random Forest. The results of the study show that SVM with a degree 2 Polynomial Kernel is the best classifier for the classification of optimal air humidity for the breeding of *Aedes aegypti* mosquitoes in Bandung City.

Keywords: Classification, Air Humidity, *Aedes aegypti*, Naïve Bayes, SVM, Random Forest

Abstrak: Salah satu penyakit yang masih menjadi perhatian di Indonesia adalah Demam Berdarah Dengue (DBD). Kota Bandung menjadi daerah dengan kasus DBD terbanyak di Indonesia pada periode Januari-April tahun 2024. Penyakit DBD disebarkan oleh nyamuk *Aedes aegypti*. Nyamuk tersebut berkembang biak dengan optimal pada kelembapan 70–90 %. Penelitian ini bertujuan untuk melakukan klasifikasi kelembapan udara optimal bagi perkembangbiakan nyamuk *Aedes aegypti*. Data yang digunakan bersumber dari BMKG. Terdapat tiga metode yang digunakan dalam pengklasifikasian yaitu Naïve Bayes, SVM, dan Random Forest. Hasil penelitian menunjukkan bahwa SVM dengan kernel polinomial derajat 2 merupakan *classifier* terbaik untuk klasifikasi kelembapan udara yang optimal bagi perkembangbiakan nyamuk *Aedes aegypti* di Kota Bandung.

Kata kunci: Klasifikasi, Kelembapan Udara, Nyamuk *Aedes aegypti*, Naïve Bayes, SVM, Random Forest

I. PENDAHULUAN

1.1. Latar Belakang

Salah satu penyakit yang masih menjadi perhatian di Indonesia adalah Demam Berdarah *Dengue* (DBD). Pada Maret 2024, terdapat hingga 16.000 kasus DBD dengan 124 kematian. Kasus tersebut naik hingga tiga kali lipat dibandingkan tahun 2023. Kota Bandung menjadi daerah dengan kasus DBD terbanyak pada periode bulan Januari-April 2024, yaitu hingga 3.468 kasus positif DBD. Pada tahun 2021, tercatat ada sebanyak 3.743 kasus DBD yang terjadi di Kota Bandung. Angka tersebut terus meningkat hingga tahun 2022, yaitu mencapai 5.205 kasus positif DBD [1].

DBD disebabkan oleh *dengue virus* (DENV). Penyakit ini merupakan penyakit endemik yang sering terjadi di wilayah tropis dan subtropis [2]. Infeksi virus DENV biasanya ditularkan oleh nyamuk *Aedes* [3]. Proses penularannya berasal dari nyamuk *Aedes aegypti* yang menggigit orang yang terinfeksi DENV, kemudian virus tersebut menyebar ke tubuh nyamuk. Nyamuk yang membawa DENV tersebut menggigit orang yang sehat, sehingga terjadi penularan dari orang yang terinfeksi ke orang yang sehat [4].

Salah satu faktor yang dapat mempengaruhi perkembangan jentik nyamuk *Aedes aegypti* adalah faktor lingkungan sekitar [5]. Faktor ini mencakup hal-hal seperti kelembapan dan suhu area tersebut. Biasanya, nyamuk *Aedes aegypti* bertelur pada suhu 20 hingga 30 derajat Celcius. Kemampuan nyamuk dalam menahan perubahan suhu berbeda-beda tergantung spesiesnya. Pada suhu 25 hingga 27 derajat Celcius, embrio sempurna terjadi pada 72 hari, dan pertumbuhan nyamuk berhenti pada saat yang sama yaitu ketika suhu turun di bawah 10 derajat Celcius atau naik di atas 40 derajat Celcius. Sementara itu,



kelembapan udara yang ideal untuk embrionasi dan kelangsungan hidup nyamuk berada diantara 70% dan 90% [6].

Banyaknya uap air di udara dinyatakan sebagai kelembapan udara [7]. Kelembapan udara dikatakan tinggi ketika udara tersebut mengandung banyak air. Jumlah air yang ada di udara disebabkan oleh uap air. Suhu mempunyai pengaruh yang signifikan terhadap banyaknya uap air yang tertahan di atmosfer. Saat suhu udara rendah, uap air untuk menjenuhkan udara hanya sedikit. Suhu ruangan dipengaruhi oleh pergerakan angin karena adanya variasi tekanan [8]. Udara yang lebih dingin akan naik menggantikan udara yang lebih ringan karena udara yang dingin ini akan menyusut saat malam hari dan memuai saat siang hari. Kelembapan udara di suatu wilayah dipengaruhi oleh temperatur, kualitas dan kuantitas penyinaran, arah angin, tekanan udara, vegetasi, dan keberadaan air tanah [7].

Penyebaran nyamuk *Wolbachia* menjadi salah satu program pengendalian kasus demam berdarah yang dilakukan oleh *World Mosquito Program* (WMP) [9]. *Wolbachia* adalah bakteri yang mampu hidup pada tubuh nyamuk *Aedes sp.*. Ketika nyamuk betina yang mengandung *Wolbachia* menghisap darah manusia yang sudah terdapat *dengue virus*, nyamuk tersebut tidak bisa lagi menularkan *dengue virus* ke manusia lain, sehingga persebaran virus akan terhenti [10]. Program ini sudah dilaksanakan di beberapa kota di Indonesia, yaitu Sleman, Bantul, Yogyakarta, Buleleng, dan Denpasar, yang mana selanjutnya akan lebih menyebar ke wilayah lain. Program ini terbukti menurunkan kasus DBD di Yogyakarta hingga 77% [9]. Kementerian kesehatan juga mengadopsi program tersebut dan menerapkan ke lima kota, seperti Kota Semarang, Bontang, Jakarta Barat, Kupang, dan Bandung. Akan tetapi, penerapan di Kota Bandung belum terbukti efektif karena titik penyebaran nyamuk *Wolbachia* ini hanya di satu kelurahan saja, sehingga Dinkes menghimbau untuk melakukan penyebaran lebih luas. Karena untuk menjadikan nyamuk *Wolbachia* efektif mengendalikan kasus demam berdarah diperlukan waktu yang cukup lama, hingga beberapa tahun [11]. Akan tetapi, biaya dari program ini mahal sehingga penentuan lokasi dan waktu sangat penting [12]. Agar efektif dan efisien, penyebaran nyamuk ini dapat mempertimbangkan dari kondisi lingkungan, seperti suhu atau kelembapan udara [13]. Dengan demikian, diperlukan informasi mengenai kapan waktu yang tepat untuk penyebaran nyamuk tersebut, yaitu dengan melakukan klasifikasi kelembapan optimal perkembangbiakan nyamuk *Aedes aegypti*.

Belum banyak penelitian yang melakukan klasifikasi tentang kelembapan udara, terlebih lagi kelembapan udara optimal perkembangbiakan nyamuk. Dengan demikian, tujuan dari penelitian ini yaitu untuk melakukan klasifikasi prediksi kelembapan udara yang optimal bagi perkembangbiakan nyamuk *Aedes aegypti* yang didasarkan pada temperatur minimum, temperatur maksimum, temperatur rata-rata, lama penyinaran matahari, kecepatan angin maksimum, arah angin saat kecepatan maksimum, kecepatan angin rata-rata, dan arah angin terbanyak. Dengan melakukan perbandingan analisis model data mining menggunakan metode Naïve Bayes, SVM, dan Random Forest sehingga diperoleh model data mining dengan tingkat akurasi yang paling baik. Dari hasil prediksi tersebut, dapat dijadikan solusi untuk menangani permasalahan yang ada, seperti penentuan waktu yang tepat penyebaran nyamuk *Wolbachia*. Selain itu, dapat digunakan sebagai pengambilan kebijakan lainnya baik bagi masyarakat, pemerintah, atau pihak terkait untuk bersama-sama mengurangi kasus DBD di Kota Bandung yang disebabkan oleh nyamuk *Aedes aegypti*, seperti penerapan 3M (Menguras, Menutup, dan Mendaur Ulang) serta *fogging* untuk membasmi nyamuk.

1.2. Penelitian Terkait

Perbandingan antara Naïve Bayes, SVM, dan Random Forest dalam memprediksi pola perkembangbiakan nyamuk untuk penyakit Dengue yang disebabkan oleh nyamuk jenis *Aedes aegypti*



merupakan aspek penting dalam pengendalian dan pencegahan penyakit. Kinerja machine learning ini dalam melakukan prediksi tersebut sangat krusial untuk intervensi tepat waktu dan manajemen penyakit yang efektif. Terbukti bahwa ketiga algoritma tersebut telah digunakan dalam beberapa penelitian untuk memprediksi wabah *Dengue* dan pola perkembangbiakan nyamuk, terutama dari segi kelembapan udara.

Metode pembelajaran mesin telah diterapkan untuk mempelajari dampak kelembapan terhadap perkembangbiakan *Aedes aegypti*. Misalnya, sebuah penelitian menggunakan algoritma pembelajaran mesin untuk menganalisis hubungan antara kelembapan dan keberadaan larva *Aedes aegypti* dalam sampel air. Hasil penelitian menunjukkan bahwa kelembapan merupakan faktor yang signifikan dalam memprediksi keberadaan larva *Aedes aegypti*, yang menunjukkan bahwa kelembapan memainkan peran penting dalam perkembangbiakan nyamuk ini [14].

Dalam penelitian lain, metode pembelajaran mesin digunakan untuk menganalisis dampak kelembapan terhadap perilaku terbang *Aedes aegypti*. Hasil penelitian menunjukkan bahwa kelembapan mempengaruhi pola terbang *Aedes aegypti*, dengan tingkat kelembapan yang lebih tinggi menyebabkan peningkatan aktivitas terbang [15].

1.3. Definisi Variabel

Temperatur minimum (T_n) adalah suhu terendah yang tercatat selama satu hari, sedangkan temperatur maksimum (T_x) adalah indeks ekstrem iklim yang mengukur suhu tertinggi yang tercatat dalam periode tertentu, biasanya dalam satu bulan. T_n digunakan untuk menghitung berbagai indeks iklim, seperti jumlah hari di mana temperatur minimum berada di bawah ambang batas tertentu, yang berguna untuk menilai kebutuhan energi untuk pemanasan bangunan serta dapat mempengaruhi aktivitas hewan seperti nyamuk [19]. Indeks T_x digunakan untuk menilai tingkat keparahan gelombang panas dan dampaknya pada berbagai sektor seperti pertanian, konsumsi energi, dan kesehatan manusia serta hewan, terutama serangga yang aktivitas kerjanya (misalnya terbang, memberi makan), tergantung akan maksimal pada suhu tertentu [20]. Temperatur rata-rata (T_{avg}) adalah metrik iklim yang mewakili suhu rata-rata selama periode tertentu, biasanya sebulan atau setahun. Metrik ini digunakan untuk menilai tren suhu secara keseluruhan dan sering digunakan dalam studi iklim untuk menganalisis perubahan pola suhu dari waktu ke waktu.

Suhu udara dapat mempengaruhi kepadatan perkembangbiakan nyamuk di suatu daerah. Peningkatan suhu dapat mempercepat perkembangbiakan nyamuk, tetapi hanya sampai pada batas tertentu. Nyamuk secara optimal dapat berkembang biak pada suhu 25-27 derajat celsius. Perkembangbiakan nyamuk akan berhenti pada saat suhu di bawah 10 derajat celsius dan di atas 40 derajat celsius. Pada suhu di bawah 10 derajat celsius, nyamuk akan lumpuh sehingga tidak dapat bergerak. Sementara pada suhu di atas 40 derajat celsius, nyamuk akan mati [21].

Lama penyinaran matahari, juga dikenal sebagai jam penyinaran matahari, adalah indikator klimatologis yang mengukur durasi waktu sinar matahari terlihat di lokasi tertentu di Bumi. Lama penyinaran matahari adalah variabel penting dalam berbagai bidang seperti meteorologi, hidrologi, dan pertanian, karena mempengaruhi jumlah radiasi matahari yang mencapai permukaan bumi [22].

Durasi sinar matahari mempengaruhi aktivitas nyamuk, terutama intensitas cahaya. Cahaya mempengaruhi pergerakan nyamuk untuk mencari makan dan mencari tempat istirahat. Nyamuk *Aedes aegypti* merupakan salah satu spesies yang mencari tempat beristirahat dalam keadaan gelap dan terlindung dari panasnya matahari. Paparan sinar matahari adalah salah satu faktor kunci yang bisa mempengaruhi perkembangbiakan nyamuk [23].



Kecepatan angin maksimum adalah kecepatan tertinggi yang dicapai angin dalam suatu periode tertentu, biasanya diukur selama kondisi cuaca ekstrem seperti badai atau topan. Pengukuran ini penting untuk menilai potensi kerusakan dan risiko yang terkait dengan angin kencang. Kecepatan angin maksimum sering diukur menggunakan anemometer, dan data ini digunakan dalam berbagai aplikasi, termasuk perencanaan konstruksi, keselamatan penerbangan, dan mitigasi bencana.

Kecepatan angin maksimum mempengaruhi aktivitas nyamuk. Kecepatan angin yang tinggi cenderung menghambat distribusi dan tingkat keberhasilan nyamuk dalam berkembang biak. Hal tersebut terjadi karena nyamuk tidak kuat dalam menghadapi angin yang kencang. Oleh karena itu, kecepatan angin menjadi faktor yang penting dalam pengendalian perkembangbiakan nyamuk dan dapat mengurangi risiko penyebaran penyakit yang dibawa [24].

Angin berperan penting dalam penyebaran nyamuk dari tempat berkembang biak mereka. Nyamuk merupakan penerbang yang lemah sehingga seringkali mengandalkan angin untuk pergerakan mereka. Arah angin dapat berpengaruh dalam ketersediaan dan kelayakan tempat berkembang biak. Arah angin memungkinkan untuk nyamuk terbang ke lingkungan yang tidak terduga. Hal tersebut dapat mempengaruhi perkembangbiakan mereka [25].

Kelembapan rata-rata, yang dilambangkan dengan RH_{avg} , adalah metrik iklim yang mewakili kelembapan relatif rata-rata dalam periode tertentu. Kemudian, kecepatan angin rata-rata adalah kecepatan rata-rata angin yang diukur selama suatu periode waktu tertentu, seperti sehari, sebulan, atau setahun.

Arah angin terbanyak adalah arah yang paling sering diambil oleh angin di suatu lokasi tertentu. Arah ini dapat bervariasi tergantung pada lokasi dan musim. Pola angin terbanyak dipengaruhi oleh pergerakan global atmosfer Bumi, termasuk angin pasat, angin barat, dan angin timur kutub. Arah angin berpengaruh terhadap bagaimana pergerakan hewan terutama serangga ketika melakukan aktivitasnya terutama dalam berburu makanan [26].

Arah angin terbanyak dapat menentukan area penyebaran nyamuk. Dengan mengetahui arah angin dominan, dapat diprediksi daerah yang kemungkinan besar akan mengalami peningkatan populasi nyamuk. Kecepatan angin rata-rata menjadi salah satu fitur utama karena angin mempengaruhi distribusi dan aktivitas nyamuk. Kecepatan angin yang rendah memungkinkan nyamuk untuk terbang dan mencari inang, sedangkan kecepatan yang tinggi dapat menghambat pergerakan nyamuk [27].

1.4. Tinjauan Literatur Metode

1. Naïve Bayes

Bayesian *classifiers* merupakan *classifiers* berbasis statistik yang dapat memprediksi peluang suatu amatan/anggota untuk masuk ke kelas tertentu. Naïve Bayes mengasumsikan bahwa atribut dari kelas independen dengan atribut lainnya. *Classifiers* ini berdasarkan pada Bayes' theorem. Alur kerja Naïve Bayes *classifiers* sebagai berikut [28]:

- a. T adalah himpunan sampel data training yang sudah memiliki label, terdapat k kelas, yaitu C_1, C_2, \dots, C_k . Tiap sampel direpresentasikan oleh vektor $X = \{x_1, x_2, \dots, x_n\}$ dan dengan atribut A_1, A_2, \dots, A_n
- b. *Classifier* akan memprediksi X untuk masuk ke suatu kelas dengan probabilitas posteriori tertinggi. X akan diprediksi masuk ke kelas C_i jika dan hanya jika $P(C_i|X) > P(C_j|X)$ untuk $1 \leq j, j \neq i$
Jadi, akan dicari kelas yang memaksimumkan $P(C_i|X)$.
- c. Karena $P(X)$ sama untuk semua kelas, maka hanya $P(X|C_i)P(C_i)$ yang perlu untuk dimaksimumkan.



- d. Untuk efisiensi dalam evaluasi $P(X|C_i)P(C_i)$, maka diperlukan asumsi independensi bersyarat.
 - e. Untuk memprediksi label kelas pada X , $P(X|C_i)P(C_i)$ dievaluasi untuk setiap kelas C_i . *Classifier* akan memprediksi label kelas X adalah C_i jika dan hanya jika kelas tersebut memaksimalkan $P(X|C_i)P(C_i)$
2. Support Vector Machine (SVM)
- SVM adalah teknik klasifikasi data linear dan non linear. SVM menggunakan pemetaan non linear yang mentransformasi data ke dalam dimensi yang lebih tinggi melalui fungsi kernel. Algoritma SVM memetakan titik data ke dalam space segingga titik data dari kategori yang berbeda dapat dibagi dengan *gap* yang jelas dan selebar mungkin [29]. Membangun *hyperplane* atau *set hyperplane* dalam ruang dimensi yang tinggi atau tak terbatas yang dapat digunakan untuk klasifikasi, regresi, atau tugas lainnya. Ada beberapa jenis kernel yang dapat mengubah data ke ruang dimensi yang lebih tinggi, di antaranya [30]:
- a. *Linear kernel*
 - b. *Polynomial function*
 - c. *Hyperbolic tangent (sigmoid)*
 - d. *Gaussian radial basis function (RBF)*
 - e. *Laplacian radial basis function*
3. Random Forest
- Random Forest adalah *classifier* yang terdiri dari beberapa *decision tree*, di mana masing-masing *decision tree* memberikan suara untuk kelas tertentu. Hasil prediksi random forest didapatkan dari suara terbanyak, yaitu kelas yang paling sering muncul atau mendapat suara dari hasil *decision tree* [31].
4. *Overfitting/Underfitting* Model
- Dalam *machine learning*, terdapat permasalahan di mana model tidak bisa menggeneralisasi *unseen data* yang diobservasi dengan baik, yaitu *overfitting*. Adanya *overfitting* membuat model menjadi *fit* pada data latih. Model menjadi menghafal pola data latih alih-alih mempelajari pola tersembunyi dari data. *Overfitting* dapat terdeteksi melalui tidak seimbangannya nilai akurasi dari prediksi data latih dan data uji, yang mana akurasi data latih jauh lebih tinggi dibandingkan data uji [32]. Sebaliknya, model juga dapat mengalami *underfitting*, yaitu saat model terlalu menggeneralisasi sehingga tidak dapat menangkap variasi/pola yang ada. *Underfitting* ditandai dengan nilai akurasi data uji yang jauh lebih tinggi dibandingkan nilai akurasi data latih [33].
5. Metode Evaluasi
- Evaluasi diperlukan untuk mengetahui seberapa tepat hasil klasifikasi yang dilakukan. Terdapat beberapa metode dalam pengujian data mining, yang paling sering digunakan antara lain nilai *accuracy*, *precision*, *recall*, dan *f-1 score* [34]. Nilai-nilai yang dicari dapat dihitung dengan rumus sebagai berikut.
- a. Akurasi
- Nilai akurasi digunakan untuk mengukur seberapa besar akurasi dari masing-masing metode dan kernel yang digunakan. Semakin besar nilai akurasi yang dihasilkan maka semakin baik metode yang digunakan.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$



Dimana TP diartikan sebagai *True Positive*, TN sebagai *True Negative*, FP sebagai *False Positive*, FN sebagai *False Negative*.

b. Presisi

Precision atau presisi merupakan ukuran perbandingan antara jumlah data yang bernilai positif dengan jumlah data benar yang bernilai positif dan data salah yang bernilai positif.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

c. Recall

Recall adalah ukuran perbandingan antara nilai jumlah data yang bernilai positif dengan jumlah data benar yang bernilai positif dan data salah yang bernilai negatif.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

d. F1-score

F-1 score dihitung dari hasil perkalian precision dan recall, kemudian dibagi dengan hasil penjumlahan precision dan recall

$$F - 1 \text{ Score} = \frac{2(Precision \times Recall)}{Precision + Recall} \quad (4)$$

II. METODE PENELITIAN

2.1. Data dan Sumber Data

Penelitian dilakukan di Kota Bandung, Jawa Barat yang memiliki kasus DBD tertinggi di Indonesia. Penelitian ini menggunakan pendekatan time-series pada rentang waktu 1 Mei 2022 - 1 Mei 2024. Data yang digunakan adalah data sekunder berupa data prakiraan cuaca harian yang didapatkan dari laman resmi Badan Meteorologi, Klimatologi, dan Geofisika (<https://dataonline.bmkg.go.id>). Data dikumpulkan dari Stasiun Geofisika Bandung. Data memiliki observasi sebanyak 732 amatan, di mana amatan tersebut berupa hari. Variabel dependen yang digunakan adalah kategori kelembapan udara optimal perkembangan nyamuk *Aedes aegypti* dengan 2 kategori, yaitu optimal dan kurang optimal. Kategori optimal merupakan kategori saat kelembapan udara berada pada angka 70–90%. Sementara kategori tidak optimal adalah kategori saat kelembapan udara berada pada angka <70% dan >90%. Kategorisasi variabel dependen dilakukan pada data kelembapan rata-rata (%). Sementara itu, variabel independen yang digunakan di antaranya:

Tabel 1. Definisi Variabel

Variabel	Nama Variabel	Definisi Operasional	Sumber Data	Satuan
RH_avg	Kelembaban udara rata-rata	Banyaknya kadar uap air yang terkandung di udara	BMKG	Persen
Tn	Temperatur minimum	Suhu udara terendah	BMKG	Celcius
Tx	Temperatur maksimum	Suhu udara tertinggi	BMKG	Celcius
Tavg	Temperatur rata-rata	Suhu udara rata-rata	BMKG	Celcius
ss	Lama penyinaran matahari	Lamanya matahari menyinari hingga permukaan bumi dalam satu hari yang diukur dalam jam	BMKG	Jam
ff_x	Kecepatan angin maksimum	Kecepatan aliran udara maksimum dari tekanan tinggi ke tekanan rendah	BMKG	m/s



Variabel	Nama Variabel	Definisi Operasional	Sumber Data	Satuan
ddd_x	Arah angin saat kecepatan maksimum	Arah dimana angin bertiup saat kecepatannya maksimum atau dikenal dengan mata angin dengan derajat 0 hingga 360 searah jarum jam	BMKG	Derajat
ff_avg	Kecepatan angin rata-rata	Kecepatan aliran udara rata-rata dari tekanan tinggi ke tekanan rendah	BMKG	m/s
ddd_car	Arah angin terbanyak	Arah datangnya angin terbanyak yang terdiri dari delapan mata angin	BMKG	Derajat

2.2. Metode yang Digunakan

Metode yang digunakan dalam penelitian ini adalah metode pada data mining berupa klasifikasi. Klasifikasi yang dilakukan di antaranya Naïve Bayes, SVM, dan Random Forest. Klasifikasi dengan SVM menggunakan beberapa kernel, seperti kernel linier, radial, polinomial derajat 2, dan sigmoid. Pemilihan model yang digunakan dalam penelitian ini didasarkan pada keunggulan dari setiap model. Model Naïve Bayes digunakan karena kesederhanaan dan efisiensinya dalam menangani dataset yang besar [16]. Selanjutnya, SVM memiliki kemampuan yang kuat dalam melakukan klasifikasi baik pada data linier maupun non-linier. SVM bekerja dengan mencari *hyperplane* yang memaksimalkan margin antara dua kelas [17]. Penggunaan berbagai kernel dalam SVM juga memungkinkan model untuk menyesuaikan diri dengan berbagai bentuk data dan memberikan fleksibilitas dalam menangkap pola yang berbeda-beda dalam data. Sementara itu, Random Forest mempunyai kekuatan dalam mengatasi overfitting dan kemampuannya dalam memberikan prediksi yang akurat melalui *ensemble learning* [18]. Setelah dilakukan pemodelan dengan ketiga algoritma tersebut, dilakukan perbandingan hasil menggunakan nilai akurasi, presisi, *recall*, dan *F1-score* untuk menentukan *classifier* terbaik.

Secara rinci, penelitian dilakukan melalui langkah-langkah berikut.

1. Pengumpulan Data

Pengumpulan data dilakukan melalui *website* resmi BMKG pada menu dataonline.bmkg.go.id dengan memilih variabel yang ditentukan dan stasiun yang diinginkan, yaitu Stasiun Geofisika Bandung.

2. Pra-Pemrosesan Data

Dalam pemrosesan data, dilakukan beberapa hal di antaranya:

- a. Penghapusan data tidak terukur
Data terukur yaitu data dengan isian 8888 atau 9999.
- b. Penghapusan *missing values*
Missing values tidak bisa dilakukan imputasi karena berupa data harian, sehingga dilakukan penghapusan.
- c. Pengategorisasian kelembaban udara
Kategorisasi berdasarkan kelembaban udara optimal untuk perkembangbiakan nyamuk *Aedes aegypti*. Terdapat 2 kategori, yaitu optimal dan kurang optimal.
- d. Perubahan tipe data
Dilakukan beberapa penyesuaian tipe data, khususnya pada data kategorik yang belum terbaca sebagai *factor*.

3. Pembagian Data menjadi Data Latih dan Data Uji

Data dibagi menjadi dua bagian dengan proporsi 80:20 yaitu menjadi data latih dan data uji.

4. Klasifikasi dengan Metode *Machine Learning*

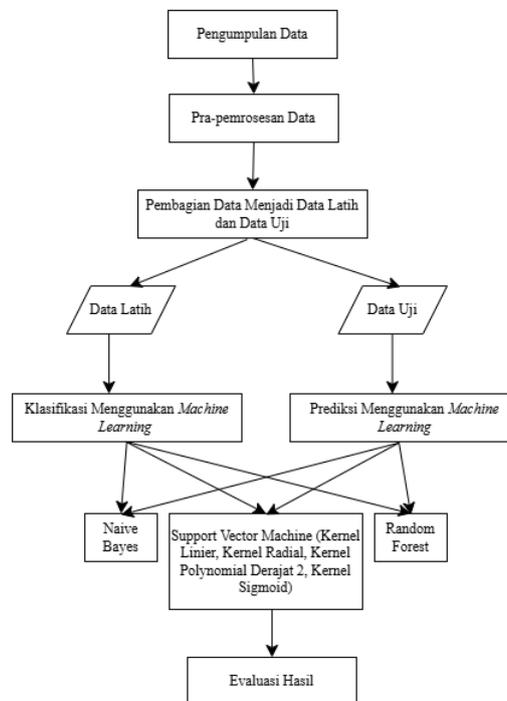
Model klasifikasi dibentuk menggunakan beberapa algoritma, seperti Naïve Bayes, SVM, dan Random Forest.

5. Prediksi dengan Metode *Machine Learning*

Model prediksi dibentuk menggunakan model hasil klasifikasi dengan beberapa algoritma, seperti Naïve Bayes, SVM, dan Random Forest.

6. Evaluasi Hasil

Evaluasi model dilakukan pada data latih dan data uji. Evaluasi yang digunakan di antaranya *accuracy*, *precision*, *recall*, dan *f-1 score*. Perbandingan akurasi data latih dan data uji untuk melihat ada tidaknya *overfitting*. Lalu, ukuran evaluasi pada data uji dilakukan untuk memilih model terbaik yang dapat digunakan.



Gambar 1. Alur Penelitian

III. HASIL DAN PEMBAHASAN

3.1. Hasil

Pengumpulan Data

Data pada penelitian ini merupakan data harian yang memiliki amatan sebanyak 732 amatan.

Pra-pemrosesan Data

1. Menghapus Data Tidak Terukur

Terdapat 49 data yang teridentifikasi sebagai data tidak terukur yang dikodekan dengan isian 8888. Karena tidak dimungkinkan melakukan imputasi, maka dilakukan penghapusan terhadap data tersebut, sehingga data yang tersisa 683 amatan.

2. Membuang Data *Missing Values*



Tabel 2. Data Missing Value

Variabel	Missing Values
Temperatur minimum (Tn)	27
Temperatur maksimum (Tx)	11
Tempetratur rata-rata (Tavg)	12
Kelembapan rata-rata (RH_avg)	12
Lama penyinaran matahari (ss)	10
Kecepatan angin maksimum (ff_x)	7
Arah angin saat kecepatan maksimum (ddd_x)	7
Kecepatan angin rata-rata (ff_avg)	7
Arah angin terbanyak (ddd_car)	7

Karena tidak dimungkinkan melakukan imputasi, maka data yang berupa missing values akan dihapus. Dengan demikian, data yang tersisa adalah 652 amatan.

3. Kategorisasi Kelembapan Udara Optimal

Tabel 3. Kategori Kelembaban Udara

Kategori Kelembapan Udara	Rentang
Optimal	70-90%
Kurang Optimal	< 70% dan > 90%

Kelembaban udara bagi perkembangbiakan nyamuk *Aedes aegyti* dikategorikan menjadi dua, yaitu kategori optimal dan kurang optimal. Kategori optimal berada direntang 70 hingga 90 persen, sedangkan kategori kurang optimal berada dibawah 70% dan diatas 90%.

4. Mengubah Tipe Data

Tabel 4. Tipe Data

Variabel	Tipe Data
Temperatur minimum (Tn)	Numerik
Temperatur maksimum (Tx)	Numerik
Tempetratur rata-rata (Tavg)	Numerik
Kelembapan rata-rata (RH_avg)	Kategorik
Lama penyinaran matahari (ss)	Numerik
Kecepatan angin maksimum (ff_x)	Numerik
Arah angin saat kecepatan maksimum (ddd_x)	Numerik
Kecepatan angin rata-rata (ff_avg)	Numerik
Arah angin terbanyak (ddd_car)	Kategorik

Dilakukan perubahan tipe data pada data yang digunakan. Variabel kelembapan rata-rata (RHavg) dan variabel arah angin terbanyak (ddd_car) menjadi tipe data kategorik. Sedangkan variabel lainnya bertipekan numerik.

Pembagian Data

Setelah dilakukan pra-pemrosesan data, selanjutnya dilakukan pembagian data menjadi data latih dan data uji. *Dataset* yang ada dibagi kedalam dua bagian dengan proporsi 80:20, yaitu data *training* atau data latih sebanyak 80% sedangkan sisanya data *testing* atau data uji 20%. Dengan demikian, data latih memiliki 521 amatan dan data uji memiliki 131 amatan.

Pengecekan *Overfitting* dan *Undefitting* Model

Tabel 5. Confusion Matrix

Algoritma	Akurasi Data Latih	Akurasi Data Uji
Naïve Bayes	0,906	0,878
SVM kernel linier	0,914	0,893
SVM kernel radial	0,917	0,877



Algoritma	Akurasi Data Latih	Akurasi Data Uji
SVM kernel polinomial derajat 2	0,889	0,916
SVM kernel sigmoid	0,917	0,878
Random Forest	1,000	0,893

Setelah dilakukan pengecekan akurasi antara data latih dan data uji, didapatkan nilai akurasi yang tidak terlalu timpang antara keduanya untuk semua model. Tidak terjadi *overfitting* model yang membuat model sangat sempurna memprediksi data train dan membuat tidak dapat memprediksi data uji dengan baik. Selain itu, tidak terjadi juga *underfitting* yang membuat model tidak dapat memprediksi data dengan baik. Karena tidak terjadi *overfitting* maupun *underfitting*, maka dapat dilakukan prediksi untuk data uji.

Hasil Prediksi dengan Model Terbaik

Prediksi dilakukan menggunakan model terbaik, yaitu model SVM dengan kernel polinomial derajat 2. Hasil yang didapatkan sebagai berikut.

Tabel 6. Hasil Prediksi dengan SVM kernel polinomial derajat 2

No	Prediksi	No	Prediksi
1	Optimal	100	Kurang Optimal
2	Optimal	101	Kurang Optimal
3	Optimal	102	Optimal
4	Optimal	103	Optimal
5	Optimal	104	Optimal
.	.	105	Optimal
.	.	106	Optimal
.	.	.	.
.	.	.	.
.	.	.	.
92	Optimal	.	.
93	Optimal	.	.
94	Kurang	.	.
95	Optimal	.	.
96	Kurang Optimal	128	Optimal
97	Kurang Optimal	129	Optimal
98	Kurang Optimal	130	Optimal
99	Kurang Optimal	131	Optimal

Dari hasil prediksi di atas, didapatkan bahwa dalam 131 hari terakhir sebelum 1 Mei 2024, kelembaban udara di Kota Bandung didominasi oleh kategori optimal. Artinya, pada 131 hari itu, nyamuk sedang dalam masa paling optimal untuk berkembang biak. Waktu tersebut merupakan waktu yang optimal untuk menerapkan berbagai program dan kebijakan, salah satunya penyebaran nyamuk *Wolbachia*. Nyamuk *Wolbachia* dalam kondisi yang kelembaban optimal dapat berkembang biak dengan baik, sehingga nyamuk *Aedes aegypti* yang berkembang adalah nyamuk yang sudah dimasuki bakteri *Wolbachia*, bukan lagi nyamuk yang menyebarkan virus *dengue*.

Confusion Matrix Hasil Prediksi

1. Naïve Bayes

Tabel 7. Confusion Matrix Naïve Bayes

	Kurang Optimal	Optimal
Kurang Optimal	15	5
Optimal	11	100



Tabel 7 menyajikan hasil prediksi dengan algoritma Naïve Bayes. Terlihat bahwa dari data yang telah diujikan, sebanyak 15 data yang berkategori kurang optimal diprediksi sebagai kategori kurang optimal dan 5 data yang berkategori kurang optimal diprediksi sebagai kategori optimal. Selanjutnya, terdapat 100 data berkategori optimal diprediksi sebagai kategori optimal dan 11 data yang berkategori optimal diprediksi sebagai kategori kurang optimal.

2. SVM Kernel Linier

Tabel 8. Confusion Matrix SVM Kernel Linier

	Kurang Optimal	Optimal
Kurang Optimal	11	9
Optimal	5	106

Tabel 8 menyajikan hasil prediksi dengan algoritma SVM kernel linier. Terlihat bahwa dari data yang telah diujikan, sebanyak 11 data yang berkategori kurang optimal diprediksi sebagai kategori kurang optimal dan 9 data yang berkategori kurang optimal diprediksi sebagai kategori optimal. Selanjutnya, terdapat 106 data berkategori optimal diprediksi sebagai kategori optimal dan 5 data yang berkategori optimal diprediksi sebagai kategori kurang optimal.

3. SVM Kernel Radial

Tabel 9. Confusion Matrix SVM Kernel Radial

	Kurang Optimal	Optimal
Kurang Optimal	10	10
Optimal	6	105

Tabel 9 menyajikan hasil prediksi dengan algoritma SVM kernel radial. Terlihat bahwa dari data yang telah diujikan, sebanyak 10 data yang berkategori kurang optimal diprediksi sebagai kategori kurang optimal dan 10 data yang berkategori kurang optimal diprediksi sebagai kategori optimal. Selanjutnya, terdapat 105 data berkategori optimal diprediksi sebagai kategori optimal dan 6 data yang berkategori optimal diprediksi sebagai kategori kurang optimal.

4. SVM Kernel Polinomial Derajat 2

Tabel 10. Confusion Matrix SVM Kernel Polinomial Derajat 2

	Kurang Optimal	Optimal
Kurang Optimal	10	10
Optimal	1	110

Tabel 10 menyajikan hasil prediksi dengan algoritma SVM kernel polinomial derajat 2. Terlihat bahwa dari data yang telah diujikan, sebanyak 10 data yang berkategori kurang optimal diprediksi sebagai kategori kurang optimal dan 10 data yang berkategori kurang optimal diprediksi sebagai kategori optimal. Selanjutnya, terdapat 110 data berkategori optimal diprediksi sebagai kategori optimal dan 1 data yang berkategori optimal diprediksi sebagai kategori kurang optimal.

5. SVM Kernel Sigmoid

Tabel 11. Confusion Matrix SVM Kernel Sigmoid

	Kurang Optimal	Optimal
Kurang Optimal	13	7
Optimal	9	102

Tabel 11 menyajikan hasil prediksi dengan algoritma SVM kernel sigmoid. Terlihat bahwa dari data yang telah diujikan, sebanyak 13 data yang berkategori kurang optimal diprediksi sebagai kategori kurang optimal dan 7 data yang berkategori kurang optimal diprediksi sebagai kategori



optimal. Selanjutnya, terdapat 102 data berkategori optimal diprediksi sebagai kategori optimal dan 9 data yang berkategori optimal diprediksi sebagai kategori kurang optimal.

6. Random Forest

Tabel 12. Confusion Matrix Random Forest

	Kurang Optimal	Optimal
Kurang Optimal	12	8
Optimal	6	105

Tabel 12 menyajikan hasil prediksi dengan algoritma Random Forest. Terlihat bahwa dari data yang telah diujikan, sebanyak 12 data yang berkategori kurang optimal diprediksi sebagai kategori kurang optimal dan 8 data yang berkategori kurang optimal diprediksi sebagai kategori optimal. Selanjutnya, terdapat 105 data berkategori optimal diprediksi sebagai kategori optimal dan 6 data yang berkategori optimal diprediksi sebagai kategori kurang optimal.

Performa Confusion Matrix

Tabel 13. Hasil Prediksi dengan SVM kernel polinomial derajat 2

Algoritma	Akurasi	Recall	Precision	F-1 Score
Naïve Bayes	0,878	0,750	0,577	0,652
SVM kernel linier	0,893	0,550	0,688	0,611
SVM kernel radial	0,878	0,500	0,625	0,556
SVM kernel polinomial derajat 2	0,916	0,500	0,909	0,645
SVM kernel sigmoid	0,878	0,650	0,591	0,619
Random Forest	0,893	0,600	0,667	0,632

Berdasarkan *confusion matrix* setiap model klasifikasi, elemen diagonal matriks memiliki nilai yang lebih besar dibandingkan elemen di luar diagonal. Elemen diagonal merepresentasikan banyak klasifikasi yang akurat untuk setiap kelas [35]. Dengan demikian, didapatkan bahwa semua model dapat mengklasifikasikan dengan akurat.

3.2. Pembahasan

Dari hasil klasifikasi dengan enam algoritma, diperoleh nilai akurasi hasil prediksi sebesar 88% pada algoritma Naïve Bayes, 89% pada algoritma SVM Kernel Linear, 88% pada SVM Kernel Radial, 92% pada algoritma SVM Kernel Polinomial dengan derajat dua, 88% pada algoritma SVM Kernel Sigmoid, dan 89% pada algoritma Random Forest. Hasil akurasi dari ketiga metode menunjukkan akurasi paling rendah sebesar 88% dan akurasi paling tinggi sebesar 92%.

Tabel 13 merupakan hasil akurasi dari algoritma yang digunakan dalam melakukan klasifikasi dan prediksi kelembapan udara optimal untuk perkembangan nyamuk *Aedes aegypti* pada Kota Bandung. Dari hasil tersebut, terlihat bahwa Algoritma SVM kernel polinomial dengan derajat dua merupakan algoritma yang paling baik dalam mengklasifikasi dataset kelembapan udara karena memiliki nilai akurasi yang paling tinggi, yaitu sebesar 92%.

SVM terpilih menjadi *classifier* terbaik karena algoritma tersebut memiliki keunggulan dapat mengolah data berdimensi tinggi tanpa terjadi penurunan performa. SVM juga dikenal sebagai salah satu *classifier* terbaik dibandingkan metode lainnya [36]. Bahkan, SVM menjadi solusi dari berbagai permasalahan klasifikasi yang banyak terjadi [37]. Selain itu, Naive Bayes tidak lebih baik dibandingkan SVM karena adanya asumsi independensi yang mana sulit diterapkan pada data yang ada [38]. Data yang digunakan dalam penelitian ini tidak memenuhi asumsi tersebut, di mana data merupakan data harian yang dapat terkait antarwaktu serta terkait antarvariabel satu sama lain. Random



Forest juga tidak lebih baik dibandingkan SVM karena algoritma tersebut cenderung bias saat terdapat variabel kategorik [39]. Penelitian ini juga menggunakan variabel bertipe kategorik dengan banyak kategori, yang mana kurang bisa ditangani dengan baik oleh Random Forest.

Kernel polinomial derajat 2 terpilih menjadi kernel terbaik dalam klasifikasi dengan SVM yang menandakan bahwa data kelembapan udara tidak bisa dilakukan pemisahan untuk klasifikasi hanya dalam satu dimensi saja. Dalam ruang satu dimensi, batas kelas hasil pemisahan tidak terlalu jelas, sehingga data tidak dapat dipisahkan secara optimal. Hal tersebut membuat data banyak yang masuk pada klasifikasi yang salah. Adanya hubungan nonlinear dari data menyebabkan perlu dilakukannya pemisahan dalam ruang dimensi yang lebih tinggi, yaitu dua dimensi.

Model klasifikasi dan prediksi tersebut dapat digunakan untuk pengambilan kebijakan, khususnya terkait pemilihan waktu untuk penyebaran nyamuk *Wolbachia* dalam rangka penurunan kasus DBD di Kota Bandung. Selain itu, pemerintah Kota Bandung juga dapat menghimbau masyarakat untuk menerapkan 3M (Menguras, Menutup, dan Mendaur Ulang) khususnya pada saat kelembapan udara di Kota Bandung sedang optimal untuk nyamuk *Aedes aegypti* berkembang. Pemerintah juga dapat melakukan *fogging* nyamuk untuk pemberantasan nyamuk yang lebih maksimal.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, didapatkan *classifier* terbaik untuk klasifikasi dan prediksi kelembapan udara yang optimal untuk perkembangbiakan nyamuk *Aedes aegypti* di Kota Bandung. Diperoleh tingkat akurasi sebesar 92%, presisi sebesar 91%, recall sebesar 50%, dan f-1 score sebesar 65%. Model tersebut dapat menjadi dasar pengambilan kebijakan terkait penentuan waktu penyebaran nyamuk *Wolbachia*, penggiatan kegiatan 3M pada masyarakat, dan pelaksanaan *fogging* nyamuk. Untuk penelitian selanjutnya, dapat dilakukan klasifikasi dan prediksi dengan variabel independen lain yang lebih relevan, khususnya terkait dengan penyakit DBD, serta algoritma *machine learning* lain yang ada.

REFERENSI

1. Dinkes Kota Bandung, “Profil Kesehatan Bandung,” *Dinas Kesehat. Kota Bandung*, p. 1, 2022, [Online]. Available: <https://dinkes.bandung.go.id/wp-content/uploads/2021/08/Versi-4-Profil-Kesehatan-Kota-Bandung-Tahun-2020.pdf>.
2. D. Kobayashi *et al.*, “Comparative analysis of the susceptibility of *Aedes aegypti* and Japanese *Aedes albopictus* to all dengue virus serotypes,” *Trop. Med. Health*, vol. 51, no. 1, 2023, doi: 10.1186/s41182-023-00553-5.
3. J. Näslund, C. Ahlm, K. Islam, M. Evander, G. Bucht, and O. W. Lwande, “Emerging Mosquito-Borne Viruses Linked to *Aedes aegypti* and *Aedes albopictus*: Global Status and Preventive Strategies,” *Vector-Borne Zoonotic Dis.*, vol. 21, no. 10, pp. 731–746, 2021, doi: 10.1089/vbz.2020.2762.
4. I. M. Sudarmaja, I. K. Swastika, L. P. E. Diarthini, I. P. D. Prasetya, and I. M. D. A. Wirawan, “Dengue virus transovarial transmission detection in *Aedes aegypti* from dengue hemorrhagic fever patients’ residences in Denpasar, Bali,” *Vet. World*, vol. 15, no. 4, pp. 1149–1153, 2022, doi: 10.14202/vetworld.2022.1149-1153.
5. N. J. Govella *et al.*, “PLOS ONE Entomological assessment of hessian fabric transfluthrin vapour emanators for protecting against outdoor-biting *Aedes aegypti* in coastal Tanzania,” pp. 1–27, 2024, doi: 10.1371/journal.pone.0299722.
6. H. Suhu *et al.*, “Keberadaan Jentik Nyamuk Penular Demam Berdarah Dengue Di Rw 01 Kelurahan Sendangguwo Semarang,” 2013.
7. K. Rahayu, E. Merita, and Permatadeny, Ary, “Tahapan eksperimen pengukuran suhu dan kelembapan pada kamar berdinding keramik dijelaskan dalam gambar diagram alir Gambar 1 .,” *J. InkoFar*, vol. 1, no. 2, pp. 5–11, 2020.
8. R. Zulfa, W. C. Lo, P. C. Cheng, M. Martini, and T. W. Chuang, “Updating the Insecticide Resistance



- Status of *Aedes aegypti* and *Aedes albopictus* in Asia: A Systematic Review and Meta-Analysis,” *Trop. Med. Infect. Dis.*, vol. 7, no. 10, 2022, doi: 10.3390/tropicalmed7100306.
9. WMP, “Indonesia World Mosquito Program,” 2023.
 10. K. M. Edenborough, H. A. Flores, C. P. Simmons, and J. E. Fraser, “Using Wolbachia to Eliminate Dengue: Will the Virus Fight Back?,” *J. Virol.*, vol. 95, no. 13, 2021, doi: 10.1128/jvi.02203-20.
 11. I. D. Velez *et al.*, “Reduced dengue incidence following city-wide wMel Wolbachia mosquito releases throughout three Colombian cities: Interrupted time series analysis and a prospective case-control study,” *PLoS Negl. Trop. Dis.*, vol. 17, no. 11 November, pp. 1–20, 2023, doi: 10.1371/journal.pntd.0011713.
 12. O. J. Brady *et al.*, “The cost-effectiveness of controlling dengue in Indonesia using wMel Wolbachia released at scale: a modelling study,” *BMC Med.*, vol. 18, no. 1, pp. 1–12, 2020, doi: 10.1186/s12916-020-01638-2.
 13. P. A. Ryan *et al.*, “Environmental factors influence the local establishment of Wolbachia in *Aedes aegypti* mosquitoes in two small communities in central Vietnam,” *Gates Open Res.*, vol. 5, no. May, 2022, doi: 10.12688/gatesopenres.13347.2.
 14. T. M. Carvajal, K. M. Viacrusis, L. F. T. Hernandez, H. T. Ho, D. M. Amalin, and K. Watanabe, “Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines,” *BMC Infect. Dis.*, vol. 18, no. 1, pp. 1–15, 2018, doi: 10.1186/s12879-018-3066-0.
 15. N. Javed, P. N. Paradkar, and A. Bhatti, “Flight behaviour monitoring and quantification of *Aedes aegypti* using convolution neural network,” *PLoS One*, vol. 18, no. 7 JULY, pp. 1–19, 2023, doi: 10.1371/journal.pone.0284819.
 16. S. Gan, S. Shao, L. Chen, L. Yu, and L. Jiang, “Adapting hidden naive bayes for text classification,” *Mathematics*, vol. 9, no. 19, pp. 1–14, 2021, doi: 10.3390/math9192378.
 17. M. Ryu and K. Lee, “Selection of support vector candidates using relative support distance for sustainability in large-scale support vector machines,” *Appl. Sci.*, vol. 10, no. 19, pp. 1–14, 2020, doi: 10.3390/app10196979.
 18. R. Natras, B. Soja, and M. Schmidt, “Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting,” *Remote Sens.*, vol. 14, no. 15, pp. 1–34, 2022, doi: 10.3390/rs14153547.
 19. C. Baril *et al.*, “The influence of weather on the population dynamics of common mosquito vector species in the Canadian Prairies,” *Parasites and Vectors*, vol. 16, no. 1, pp. 1–14, 2023, doi: 10.1186/s13071-023-05760-x.
 20. J. M. Reinhold, C. R. Lazzari, and C. Lahondère, “Effects of the environmental temperature on *Aedes aegypti* and *Aedes albopictus* mosquitoes: A review,” *Insects*, vol. 9, no. 4, 2018, doi: 10.3390/insects9040158.
 21. O. R. Pinontoan, O. J. Sumampouw, J. H. V. Ticoalu, J. E. Nelwan, E. C. Musa, and J. Sekeeon, “The variability of temperature, rainfall, humidity and prevalence of dengue fever in Manado City,” *Bali Med. J.*, vol. 11, no. 1, pp. 81–86, 2022, doi: 10.15562/bmj.v11i1.2722.
 22. B. Wu, S. Liu, W. Zhu, M. Yu, N. Yan, and Q. Xing, “A method to estimate sunshine duration using cloud classification data from a geostationary meteorological satellite (FY-2D) over the Heihe River Basin,” *Sensors (Switzerland)*, vol. 16, no. 11, 2016, doi: 10.3390/s16111859.
 23. A. Aunillah and T. Kesetyaningsih, “Relationship Between Rainy Day and Sunshine Duration with Dengue Hemorrhagic Fever Incidence in City of Yogyakarta,” *Proc. Univ. Muhammadiyah Yogyakarta Undergrad. Conf.*, vol. 2, no. 2, pp. 16–23, 2022, doi: 10.18196/umygrace.v2i2.420.
 24. A. Sutriyawan *et al.*, “Analysis of Temperature, Humidity, Rainfall, and Wind Velocity on Dengue Hemorrhagic Fever in Bandung Municipality,” *Russ. J. Infect. Immun.*, vol. 14, no. 1, pp. 155–162, 2024, doi: 10.15789/2220-7619-AOT-2110.
 25. X. Ju *et al.*, “How air pollution altered the association of meteorological exposures and the incidence of dengue fever,” *Environ. Res. Lett.*, vol. 17, no. 12, 2022, doi: 10.1088/1748-9326/aca59f.
 26. M. W. Hayward, G. J. Hayward, and G. I. H. Kerley, “Space use of ungulate prey relative to lions is affected by prey species and predator behavior but not wind direction,” *Ecol. Res.*, vol. 38, no. 5, pp. 639–650, 2023, doi: 10.1111/1440-1703.12397.
 27. A. C. A. Dias, M. M. S. Rodrigues, and A. A. Silva, “Effect of acute and chronic exposure to ammonia on different larval instars of *Anopheles darlingi* (Diptera: Culicidae),” *J. Vector Ecol.*, vol. 44, no. 1, pp. 112–118, 2019, doi: 10.1111/jvec.12335.
 28. Y. Karaca and C. Cattani, “7. Naive Bayesian classifier,” *Comput. Methods Data Anal.*, pp. 229–250,



- 2018, doi: 10.1515/9783110496369-007.
29. M. Sun, “Support Vector Machine Models for Classification,” *Encycl. Bus. Anal. Optim.*, pp. 2395–2409, 2014, doi: 10.4018/978-1-4666-5202-6.ch215.
 30. S. Amarappa and S. V Sathyanarayana, “Data classification using Support vector Machine (SVM), a simplified approach,” *Int. J. Electron. Comput. Sci. Eng.*, vol. 3, pp. 435–445, 2011, [Online]. Available: www.ijecse.org.
 31. J. R. Thompson and B. L. Licklider, “Visualizing Urban forestry: Using concept maps to assess student performance in a learning-centered classroom,” *J. For.*, vol. 109, no. 7, pp. 402–408, 2011, doi: 10.1093/jof/109.7.402.
 32. X. Ying, “An Overview of Overfitting and its Solutions,” *J. Phys. Conf. Ser.*, vol. 1168, no. 2, 2019, doi: 10.1088/1742-6596/1168/2/022022.
 33. H. K. Jabbar and R. Z. Khan, “Methods to Avoid Over-Fitting and Under-Fitting in Supervised Machine Learning (Comparative Study),” pp. 163–172, 2015, doi: 10.3850/978-981-09-5247-1_017.
 34. N. Arifin, U. Enri, and N. Sulistiyowati, “Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 2, p. 129, 2021, doi: 10.30998/string.v6i2.10133.
 35. D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, “Multi-label Classifier Performance Evaluation with Confusion Matrix,” pp. 01–14, 2020, doi: 10.5121/csit.2020.100801.
 36. S. Taheri and M. Mammadov, “Learning the naive bayes classifier with optimization models,” *Int. J. Appl. Math. Comput. Sci.*, vol. 23, no. 4, pp. 787–795, 2013, doi: 10.2478/amcs-2013-0059.
 37. O. Karal, “Performance comparison of different kernel functions in SVM for different k value in k-fold cross-validation,” *Proc. - 2020 Innov. Intell. Syst. Appl. Conf. ASYU 2020*, pp. 0–4, 2020, doi: 10.1109/ASYU50717.2020.9259880.
 38. F. Sodik and I. Kharisudin, “Analisis Sentimen dengan SVM , NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter,” *Prisma*, vol. 4, pp. 628–634, 2021.
 39. T. Zhu, “Analysis on the applicability of the random forest,” *J. Phys. Conf. Ser.*, vol. 1607, no. 1, 2020, doi: 10.1088/1742-6596/1607/1/012123.