



Komparasi Algoritma *Machine Learning* dalam Klasifikasi *El Niño-Southern Oscillation*

Bintana Tajmala¹, Ratih Restiani², Mutiara Nur Tsani Helfiana³, Robert Kurniawan⁴

^{1,2,3}Program Studi Statistika, Politeknik Statistika STIS
⁴Program Studi Komputasi Statistika, Politeknik Statistika STIS

¹ 212111957@stis.ac.id

³ 212112231@stis.ac.id

⁴ robertk@stis.ac.id

Corresponding author email: 212112308@stis.ac.id

Abstract: Indonesia's geographical location in the equatorial region makes Indonesia have a tropical climate and the main livelihood of the people is in the agricultural sector. Climate change is the biggest threat to Indonesia's agricultural production as an agricultural country. ENSO is an indicator that can be used as a reference in viewing climate change. Therefore, it is necessary to classify what factors can influence climate change. Classification is carried out using machine learning methods with three algorithms, namely KNN, SVM, and Naive Bayes. Classification results with the highest accuracy of 85.23% were obtained using the SVM algorithm.

Keywords: classification, ENSO, KNN, SVM, Naive Bayes

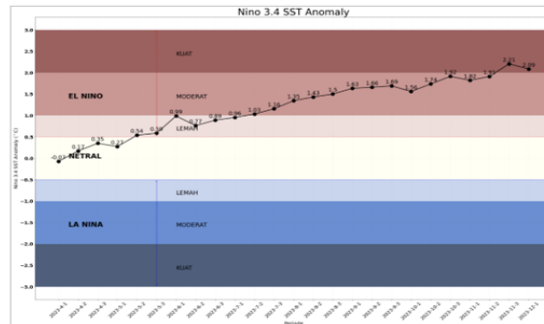
Abstrak: Letak geografis Indonesia yang berada di daerah khatulistiwa menyebabkan Indonesia beriklim tropis dan mata pencaharian utama masyarakatnya berada di sektor pertanian. Perubahan iklim merupakan ancaman terbesar dalam produksi pertanian Indonesia sebagai negara agraris. ENSO merupakan indikator yang dapat dijadikan sebagai acuan dalam melihat perubahan iklim. Dengan begitu perlu dilakukan klasifikasi faktor apa saja yang dapat memengaruhi perubahan iklim. Klasifikasi dilakukan dengan memanfaatkan metode machine learning dengan tiga algoritmanya, yaitu KNN, SVM, dan Naive Bayes. Hasil klasifikasi dengan akurasi tertinggi sebesar 85,23% didapatkan dengan menggunakan algoritma SVM.

Kata kunci: klasifikasi, ENSO, KNN, SVM, Naive Bayes

I. PENDAHULUAN

Letak geografis Indonesia berada di bawah garis khatulistiwa menyebabkan Indonesia beriklim tropis. Terletak diantara benua Australia dan Asia, antara samudra Pasifik dan Hindia, serta terdiri dari pulau dan kepulauan menyebabkan Indonesia rawan mengalami fenomena perubahan iklim [1]. Menurut NASA perubahan iklim merupakan perubahan pola cuaca rata-rata yang menentukan iklim lokal, regional, maupun global di bumi secara jangka panjang, dimana perubahan-perubahan ini dapat memberikan dampak yang luas. Indonesia merupakan negara dengan iklim tropis, sehingga sebagian besar penduduknya bekerja disektor pertanian. Tercatat pada Februari tahun 2022 produksi beras di Indonesia menduduki posisi keempat dunia [2]. Berdasarkan data Badan Pusat Statistik, pada tahun 2023 produksi Gabah Kering Giling (GKG) mencapai 53,63 juta ton. Angka ini mengalami penurunan dari yang sebelumnya 54,75 juta ton pada tahun 2022. Perubahan iklim merupakan ancaman bagi negara-negara agraris, seperti Indonesia [3].

Perubahan iklim ini dapat dilihat dari perubahan El Niño-Southern Oscillation (ENSO). Berdasarkan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), ENSO merupakan anomali dari suhu permukaan laut di Samudera Pasifik di Pantai barat Ekuador dan Peru yang lebih tinggi daripada rata-rata normalnya. Iklim di Samudera Pasifik ada tiga jenis, yaitu fase netral, fase El Niño, dan fase La Niña. Produktivitas padi dipengaruhi oleh kondisi saat fase El Niño terjadi. Hal ini karena pada fase ini suhu permukaan laut perairan Indonesia yang biasanya hangat menjadi lebih dingin, sehingga menyebabkan kurangnya curah hujan di Indonesia [4]. Pendekatan anomali yang digunakan Indonesia dengan mempertimbangkan letak geografisnya adalah dengan menggunakan NINO 3.4.



Gambar 1. Grafik Perubahan NINO 3.4 Tahun 2023

Terlihat pada gambar 1 di atas, bahwa sebagian besar kondisi yang dialami Indonesia pada tahun 2023 adalah fase El Niño. Berdasarkan uraian di atas, disebutkan juga bahwa fase ini sangat berpengaruh pada produktivitas padi di Indonesia. Dengan demikian perlu dilakukan analisis terkait pengelompokan fase ENSO yang terjadi sehingga dapat dilakukan inisiasi rencana dalam menghadapi fase yang terjadi.

Klasifikasi merupakan salah satu metode yang digunakan dalam *machine learning* yang dapat diterapkan dalam bidang meteorologi dan klimatologi [5]. Menurut penelitian yang telah dilakukan oleh [6], K-Nearest Neighbour (KNN) merupakan algoritma yang memiliki performa terbaik dengan akurasi sebesar 0,96 dibandingkan Neural Network, Random Forest, dan SVM untuk mengklasifikasikan ENSO berdasarkan *sea surface temperature* (SST) di wilayah tropis Samudera Pasifik. Berbeda dengan penelitian oleh [7] yang menggunakan data harian global near-surface (1000 hPa) Air Temperature (SAT) dari National Oceanic and Atmospheric Administration (NOAA). Hasil klasifikasi ENSO terbaik di wilayah tropis Samudera Pasifik dihasilkan oleh metode *Random Forest* dibandingkan SVM, Naïve Bayes, KNN-3, dan metode klasifikasi lain karena memiliki nilai evaluasi paling tinggi, yaitu akurasi sebesar 0,78, nilai presisi 0,77, F1 score 0,76, nilai *recall* 0,78 dan nilai AUC ROC sebesar 0,79.

Beberapa penelitian sebelumnya telah melakukan klasifikasi fase ENSO. Namun, masih sedikit penelitian yang mengklasifikasikan fase ENSO di Indonesia. Padahal penentuan fase ENSO di wilayah Indonesia sangat penting untuk sektor pertanian. Oleh karena itu, diperlukan penentuan model yang paling akurat dan efektif dalam mengklasifikasikan fase ENSO. Penelitian ini bertujuan untuk membandingkan kinerja beberapa algoritma *machine learning* dalam melakukan klasifikasi fase ENSO. Peneliti mengevaluasi dan membandingkan algoritma *K-Nearest Neighbours*, *Support Vector Machine*, dan *Naïve Bayes*. Adapun metrik evaluasi yang digunakan adalah akurasi, presisi, *recall*, F-1 score, *sensitivity*, *specificity*, dan nilai Kappa. Data yang digunakan merupakan data yang mencakup berbagai parameter meteorologi dan oceanografi yang relevan dan diambil dari sumber yang terpercaya. Hasil dari penelitian ini diharapkan dapat menjadi acuan dasar dalam melakukan prediksi fase ENSO dengan menerapkan model *machine learning* yang optimal.

II. METODE PENELITIAN

2.1 Metode

Berdasarkan kajian literatur yang ada, penelitian ini juga melakukan analisis kuantitatif dengan menggunakan beberapa metode *Machine Learning*. Metode ini merupakan salah satu ilmu komputasi statistik untuk melakukan prediksi dan optimalisasi dengan bantuan kecerdasan buatan [8]. Biasanya *machine learning* digunakan pada data yang besar atau *big data*. Ada beberapa logaritma dalam metode

ini, adapun dalam penelitian ini menggunakan logaritma K-Nearest Neighbours (KNN), Support Vector Machines (SVM), dan Naïve Bayes.

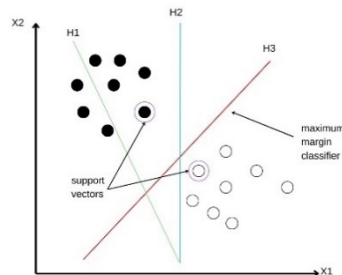
2.1.1. Algoritma KNN

Algoritma KNN merupakan salah satu metode klasifikasi non parametrik yang sederhana tetapi efektif pada banyak kasus [9]. Konsep utamanya adalah untuk menentukan kategori suatu *query* tidak hanya melalui nilai tetangga yang terdekat pada dataset, tetapi juga mempertimbangkan kategori tetangga lain sebanyak *k* terdekat atau dengan kata lain metode ini bekerja dengan mencocokkan jarak tertimbang yang berbasis vektor [10]. Pengambilan keputusannya sangatlah sederhana, yaitu sampel yang akan diuji sama dengan kategori sampel yang paling dekat dengannya. Jika data *test* dan matriks jarak tidak diubah, maka hasil keputusannya telah ditentukan secara unik untuk setiap *instance* yang akan diuji [11].

2.1.2. Algoritma SVM

Algoritma ini diperkenalkan oleh Vapnik sebagai salah satu algoritma dalam *machine learning* yang menggunakan teori statistik. Support Vector Machine (SVM) adalah metode terbaru yang paling banyak digunakan untuk melakukan klasifikasi dan prediksi dalam data [8]. SVM menjelaskan hubungan antara variabel dependen dan variabel independen dengan menemukan *hyperplane* maksimal yang bertindak sebagai margin antara dua kelas data.

Margin yang digunakan ditarik sedemikian rupa sehingga menghasilkan jarak antara margin dan kelas maksimal dalam ruang yang berdimensi *n* sehingga dapat membantu klasifikasi titik data dan meminimalkan kesalahan dalam klasifikasi. Terdapat juga *support vectors* yang menjadi titik data paling dekat dengan permukaan keputusan. Adapun proses pengklasifikasian dapat diilustrasikan dalam gambar berikut.



Gambar 2. Ilustrasi Klasifikasi dengan SVM [8]

2.1.3. Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah salah satu algoritma dalam *machine learning* yang menggunakan peluang bersyarat untuk menentukan kemungkinan suatu objek termasuk dalam kelas tertentu berdasarkan Teorema Bayes dibawah ini.

$$P\left(\frac{c}{x}\right) = \frac{P\left(\frac{x}{c}\right)P(c)}{P(x)} \quad (7)$$

Dimana $P\left(\frac{c}{x}\right)$ adalah *posterior probability*, $P\left(\frac{x}{c}\right)$ adalah fungsi *likelihood*, $P(c)$ adalah probabilitas kelas sebelumnya, dan $P(x)$ adalah probabilitas prediktor sebelumnya

Disebut naïve karena beroperasi dengan mengasumsikan bahwa fitur yang digunakan dalam pengklasifikasian tidak bergantung satu sama lain, yang berarti bahwa variabel satu dengan yang lain tidak berkaitan [8]. Naïve Bayes mengelompokkan dengan menetapkan *posterior probability* ke suatu kelas berdasarkan probabilitas dan kemungkinan sebelumnya untuk data pelatihan yang digunakan [8].



Pada dasarnya algoritma ini melanggar asumsi independensi, tetapi sebagian besar pengklasifikasian dengan Naïve Bayes digunakan dalam klasifikasi biner dan multikelas. Akan tetapi, kinerja Naïve Bayes akan melambat ketika dilakukan pemindaian ulang seluruh kumpulan data untuk setiap operasi klasifikasi baru. Algoritma ini juga sangat efektif digunakan untuk memodelkan data yang mengandung noise.

Dalam melakukan evaluasi algoritma, digunakan confusion matrix sebagai penentu algoritma mana yang paling baik digunakan untuk mengklasifikasikan ENSO. *Confussion matrix* sendiri memiliki empat kondisi yang dapat dilihat melalui tabel dibawah ini.

Tabel 1. *Confusion Matrix*

| Predicted Values | Actual Values | |
|------------------|----------------|----------------|
| | Positive | Negative |
| Positive | True Positive | False Positive |
| Negative | False Negative | True Negative |

Dalam confusion matrix ini juga terdapat empat jenis parameter evaluasi, yaitu:

1. Akurasi, yaitu total dari seberapa sering model mengklasifikasikan dengan benar. Adapun rumus dalam menghitung akurasi adalah

$$akurasi = \frac{TP+TN}{TP+FP+FN+TN} \quad (8)$$

2. Presisi, yaitu seberapa sering model dapat memprediksi benar positif. Adapun rumus dalam menghitungnya adalah

$$presisi = \frac{TP}{TP+FP} \quad (9)$$

3. Recall, merupakan seberapa sering model memprediksi positif ketika kelas aktualnya positif. Rumus dalam menghitungnya adalah sebagai berikut

$$recall = \frac{TP}{TP+FN} \quad (10)$$

4. F1-Score, ialah rata - rata harmonik dari presisi dan recall. Penghitungannya adalah sebagai berikut

$$F1 = \frac{2(recall+presisi)}{(recall+presisi)} \quad (11)$$

5. Sensitivitas, merupakan ukuran untuk mengukur kemampuan model dalam mendeteksi positif sebenarnya (*True Positive rate*), atau dengan rumus sebagai berikut

$$sensitivitas = \frac{TP}{TP+FN} \quad (12)$$

6. Spesifisitas, ialah kemampuan model dalam mendeteksi negatif sebenarnya atau *True Negative Rate*. Adapun rumusnya adalah sebagai berikut

$$spesifisitas = \frac{TN}{TN+FP} \quad (13)$$

7. Kappa (Cohen’s Kappa), ialah pengukuran untuk mengukur tingkat kesepakatan antara dua pengamat atau metode klasifikasi dengan mengoreksi kemungkinan kesepakatan terjadi secara kebetulan. Kappa menghitung kesepakatan yang diharapkan secara kebetulan. Adapun rumusnya adalah sebagai berikut

$$K = \frac{P_o - P_c}{1 - P_c} \quad (14)$$

dimana

$$P_o = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

$$P_c = \left(\frac{(TP+FP) \cdot (TP+FN) + (TN+FN) \cdot (TN+FP)}{(TP+TN+FP+FN)^2} \right) \quad (16)$$

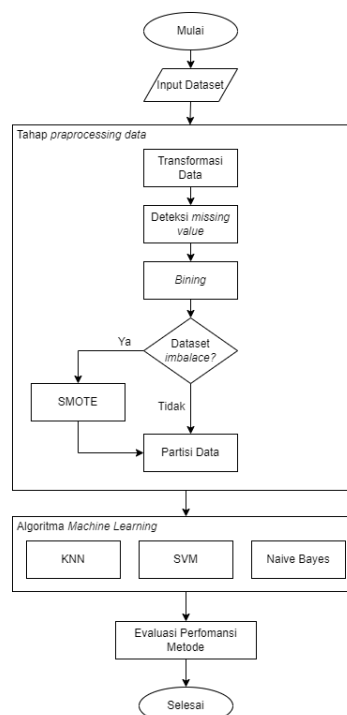
2.2 Data

Data yang digunakan pada penelitian ini merupakan data runtun waktu bulanan dari tahun 1950 sampai 2023. Variabel independen yang digunakan merupakan NINO 3.4 sebagai indikator yang digunakan sebagai penanda fenomena El Niño atau La Niña di kawasan garis khatulistiwa. Variabel-variabel dependen yang digunakan dalam penelitian ini adalah sebagai berikut:

Tabel 2. Variabel Dependen

| Variabel | Definisi Operasional | Penelitian Terkait | Sumber | Satuan |
|-----------------|--|--------------------|--------------------------------|-------------------|
| Kecepatan Angin | Pergerakan udara secara horizontal pada ketinggian 10meter diatas permukaan bumi | [12], [13] | Copernicus Climate Data Source | m s ⁻¹ |
| SST | Suhu air laut yang dekat dengan permukaan air | [13]–[15] | Copernicus Climate Data Source | K |
| OLR | Radiasi termal atau gelombang panjang yang dipancarkan ke ruang angkasa diatas atmosfer | [16]–[18] | Copernicus Climate Data Source | W m ⁻² |
| MSL | Tekanan vertikal atmosfer di permukaan bumi yang disesuaikan dengan rata-rata ketinggian permukaan air laut | [19], [20] | Copernicus Climate Data Source | Pa |
| SOI | Indikator pengembangan dan intensitas fenomena El Nino atau La Nina di Samudera Pasifik yang dihitung melalui perbedaan tekanan antara Tahiti dan Darwin | [21] | Bureau of Meteorology | |

2.3 Tahapan Penelitian



Gambar 3. Flowchart tahapan penelitian

Gambar 3., menunjukkan alur tahapan penelitian yang dilakukan. Sebelum melakukan analisis menggunakan algoritma *machine learning*, perlu dilakukan tahapan *pre-processing* data yang meliputi transformasi data, pendeteksian *missing value*, proses *binning*, dan penanganan dataset yang *imbalance* dengan SMOTE dan partisi data menjadi data *training* dan data *testing* berdasarkan dataset yang sudah *balance*. Sedangkan metode klasifikasi yang digunakan adalah KNN, SVM, dan Naive Bayes karena populer digunakan dalam menangani berbagai jenis data, masalah klasifikasi, dan aplikasi.

III. HASIL DAN PEMBAHASAN

Kejadian El Niño dan La Niña dapat dilihat dari beberapa indikator seperti NINO 3.4 dan SOI [22]. BMKG menentukan fenomena El Niño dan La Niña lewat pertimbangan indeks NINO 3.4 SST dan SOI [23]. Indonesia termasuk ke dalam daerah yang teramati lewat indeks NINO 3.4 SST. Hal ini karena Indonesia berada di wilayah Samudra Pasifik Timur dan Tengah dengan koordinat 50° LU - 50° LS dan 170° BB – 120° BB. Berikut ditampilkan klasifikasi fenomena El Niño dan La Niña menurut indeks yang digunakan:

Tabel 3. Klasifikasi NINO 3.4 *anomaly* SST

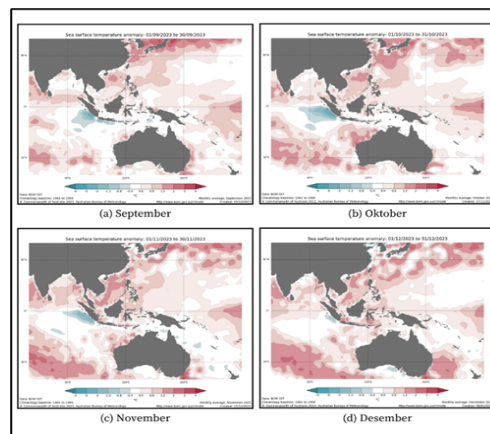
| No | NINO 3.4 SST | Fenomena |
|----|---------------|----------|
| 1 | > 0,5 | El Niño |
| 2 | -0,5 s.d. 0,5 | Netral |
| 3 | < -0,5 | La Niña |

Pada penelitian ini, hanya digunakan dua klasifikasi, yaitu fenomena El Niño dan La Niña. Apabila nilai indeks NINO 3.4 SST dalam suatu waktu dikatakan netral, dilakukan pendekatan apabila nilainya negatif maka akan terklasifikasi La Niña dan apabila nilainya positif akan terklasifikasi El Niño.

Tabel 4. Klasifikasi SOI

| No | SOI | Fenomena |
|----|-----------|----------|
| 1 | < -7 | El Niño |
| 2 | -7 s.d. 7 | Netral |
| 3 | > 7 | La Niña |

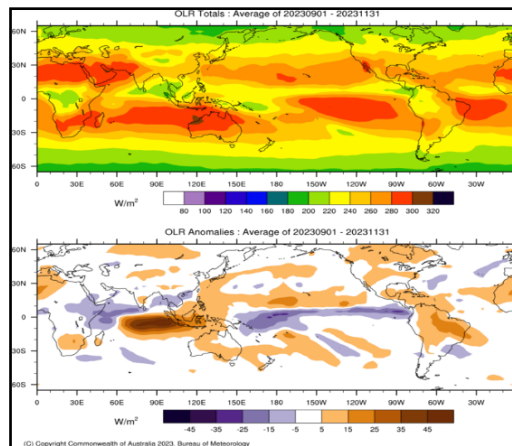
Klasifikasi fenomena El Niño dan La Niña juga dapat dilakukan melalui SOI. Pada penelitian ini, SOI menjadi variabel independen. Kategorisasi pada variabel SOI tetap dijadikan tiga kategori.



Gambar 4. Peta anomaly SST

Sementara itu pada gambar 4, dapat dilihat gambaran umum mengenai kondisi anomali suhu permukaan air laut di daerah asia bagian pasifik tropis. Pada bulan september wilayah di sekitar

Sumatera dan Jawa memiliki anomali suhu yang cukup rendah. Kemudian, pada bulan oktober hingga desember anomali suhu permukaan air laut disekitar Pulau Jawa mulai menghangat. Nilai anomali suhu permukaan air laut di timur pulau sumatera juga perlahan menghangat tetapi tidak setinggi suhu di sekitar Pulau Jawa. Kondisi anomali positif ini merupakan indikasi adanya fenomena El Niño [24]. Dimana apabila suatu daerah mengalami anomali El Niño akan menyebabkan penurunan curah hujan pada daerah tersebut [25]. Sementara itu, Pulau Sumatera memiliki anomali suhu negatif, hal ini diakibatkan karena adanya pergerakan awan evaporasi dari wilayah Samudera Hindia sehingga menyebabkan hujan jatuh di Pulau Sumatera. Hal ini sesuai dengan prediksi musim hujan yang dirilis BMKG, Pulau Sumatera telah mengalami musim hujan lebih dahulu dibandingkan Pulau Jawa dan daerah lainnya [26]. Pulau Jawa dan daerah lainnya cenderung memiliki anomali suhu permukaan air laut yang lebih hangat. Dominasi wilayah dengan anomali suhu hangat ini menjadi pengaruh kuat Indonesia memiliki indeks NINO 3.4 SST yang menunjukkan kondisi El Niño moderat.



Gambar 5. Peta Nilai Rata-rata OLR dan Anomali OLR

Berdasarkan gambar 5., terjadi anomali OLR dengan indeks positif di wilayah Indonesia. Nilainya berkisar antara 230-270 Wm^{-2} . Anomali OLR positif seringkali berkaitan dengan kondisi atmosfer yang lebih cerah dan kering. Hal ini disebabkan oleh kurangnya awan yang menghalangi radiasi inframerah dari permukaan Bumi. Dalam konteks skala global, anomali OLR positif di daerah tropis seringkali menjadi indikasi dari fenomena El Niño.

Klasifikasi El Niño dan La Niña memainkan peran krusial dalam membantu masyarakat global menghadapi tantangan yang ditimbulkan oleh perubahan iklim dan cuaca. Oleh karena itu, dalam penelitian ini dilakukan klasifikasi fenomena tersebut berdasarkan beberapa variabel terkait, menggunakan metode machine learning. Secara khusus, teknik *machine learning ensemble* dan *hybrid* telah memberikan hasil yang menjanjikan. Penelitian sebelumnya menggunakan metode KNN, Neural Network, Random Forest, dan Final Ensemble [6].

3.1 Tahap Pre-processing

Terdapat 5 variabel dan 888 amatan yang digunakan dalam penelitian ini. Variabel SOI diambil dengan tipe data numerik. Namun, dilakukan kategorisasi berdasarkan tabel 4. Oleh karena itu, tipe datanya berubah menjadi kategorik. Pada pengecekan *missing value*, hasilnya menunjukkan tidak ada amatan yang hilang pada data. Tahap selanjutnya adalah *binning*, di mana variabel NINO 3.4 akan dikategorikan menjadi dua kelompok. Nilai negatif akan diklasifikasikan sebagai La Niña dengan



kategori 1, sementara nilai positif akan diklasifikasikan sebagai El Niño dengan kategori 2. Berikut ditampilkan data dari setiap variabel yang digunakan.

| | Tahun | Bulan | Kecepatan Angin (m/s) | SST (K) | OLR (W/m ²) | MSL (Pa) | SOI_KAT | NINO_KAT2 |
|-----|-------|-----------|-----------------------|---------|-------------------------|----------|---------|-----------|
| 1 | 1950 | Januari | 5.0 | 301.0 | -249.9 | 100928.1 | 2 | 1 |
| 2 | 1950 | Februari | 5.1 | 301.1 | -240.4 | 100951.1 | 3 | 1 |
| 3 | 1950 | Maret | 4.3 | 301.4 | -235.2 | 100943.7 | 3 | 1 |
| 4 | 1950 | April | 3.6 | 302.1 | -232.9 | 100831.4 | 3 | 1 |
| 5 | 1950 | Mei | 3.3 | 302.7 | -246.8 | 100842.3 | 3 | 1 |
| 884 | 2023 | Agustus | 4.5 | 302.7 | -224.5 | 101045.9 | 1 | 2 |
| 885 | 2023 | September | 4.1 | 302.7 | -218.0 | 100990.1 | 1 | 2 |
| 886 | 2023 | Oktober | 3.2 | 303.0 | -229.9 | 101090.3 | 2 | 2 |
| 887 | 2023 | November | 3.8 | 303.0 | -222.8 | 100983.8 | 1 | 2 |
| 888 | 2023 | Desember | 5.0 | 302.5 | -241.6 | 101023.1 | 2 | 2 |

Gambar 6. Data setelah proses *binning*

Berdasarkan gambar 6., ditampilkan 5 data teratas dan 5 data terbawah pada data yang telah melalui tahap *binning*. Variabel SOI dan NINO 3.4 telah berubah menjadi kategorik sebagaimana hasil dari proses *binning*. Kemudian, dilakukan cek proporsi pada kategori variabel kejadian El Niño dan La Niña. Proporsi awal menunjukkan hasil belum seimbang (*imbalance*). Oleh karena itu, metode SMOTE digunakan untuk mengatasi data yang *imbalance*.

3.2 Pemilihan Rasio Perbandingan Sampel

Berikut disajikan hasil pembentukan model dengan prediksi data *testing* berdasarkan rasio perbandingan sampel dan tiap metode klasifikasi.

Tabel 5. Akurasi Perbandingan Rasio Sampel

| Rasio Perbandingan | Akurasi Model | | |
|--------------------|---------------|-------|-------------|
| | KNN | SVM | Naïve Bayes |
| 70 : 30 | 60,75 | 77,74 | 74,34 |
| 80 : 20 | 69,89 | 76,14 | 72,16 |
| 90 : 10 | 68,18 | 85,23 | 78,41 |

Berdasarkan hasil pada tabel 5., rasio perbandingan sampel 90% data *training* dan 10% data *testing* menghasilkan nilai akurasi tertinggi pada metode SVM dan Naive-bayes. Oleh karena itu, rasio perbandingan ini digunakan untuk penilaian performa tiap metode.

3.3 Hasil Confusion Matrix

Tabel 6. Confusion Matrix

| Data | Metode | Reference | | Prediction |
|-------|-------------|-------------|-------------|------------|
| | | La Niña (1) | El Niño (2) | |
| Train | KNN | 300 | 108 | 1 |
| | | 99 | 292 | 2 |
| | SVM | 321 | 82 | 1 |
| | | 78 | 318 | 2 |
| | Naïve Bayes | 327 | 119 | 1 |
| | | 72 | 281 | 2 |
| Test | KNN | 29 | 13 | 1 |
| | | 15 | 31 | 2 |
| | SVM | 37 | 6 | 1 |
| | | 7 | 38 | 2 |
| | Naïve Bayes | 39 | 14 | 1 |
| | | 5 | 30 | 2 |

Terdapat banyak informasi yang ditampilkan melalui *confusion matrix*. Melalui tabel *confusion matrix* dapat dihitung parameter evaluasi, seperti akurasi, presisi, *recall*, *F-1 score*, *sensitivity*, dan *specificity*



3.4 Hasil Evaluasi Performasi tiap Metode

Tabel 7. Perbandingan Parameter Evaluasi

| Data | Metode | Akurasi | Presisi | Recall | F-1 Score |
|-------|-------------|---------|---------|--------|-----------|
| Train | KNN | 74,09 | 73,53 | 75,19 | 0,74 |
| | SVM | 79,97 | 79,65 | 80,45 | 0,80 |
| | Naïve Bayes | 76,10 | 73,31 | 81,95 | 0,77 |
| Test | KNN | 68,18 | 69,04 | 65,91 | 0,67 |
| | SVM | 85,23 | 86,04 | 84,09 | 0,85 |
| | Naïve Bayes | 78,41 | 73,58 | 88,63 | 0,80 |

Terdapat beberapa parameter yang digunakan untuk evaluasi model. Pada Tabel 7., disajikan hasil evaluasi performa tiap metode berdasarkan data prediksi yang digunakan. Apabila identifikasi prediksi menggunakan data *training*, metode SVM secara konsisten memiliki nilai tertinggi pada parameter evaluasi akurasi, presisi, dan *F-1 score*. Kemudian, disusul oleh metode Naive-bayes dan setelahnya metode KNN. Namun, pada ukuran *recall*, metode Naive memiliki nilai *recall* yang lebih besar dibandingkan metode KNN dan SVM.

Berdasarkan Tabel 7., jika identifikasi prediksi menggunakan data *testing*, dapat diketahui bahwa pada ukuran akurasi, presisi dan *F-1 score* metode SVM memiliki nilai tertinggi dan disusul oleh metode Naive-bayes lalu KNN. Nilai akurasi pada metode SVM mencapai 85,23%. Kemudian, disusul oleh metode Naive-bayes dengan akurasi sebesar 78,41% dan metode KNN sebesar 68,18%. Hal ini berarti, metode SVM dapat memprediksi kejadian El Niño dan La Niña dengan tepat per seluruh kejadian. Kemudian, pada ukuran presisi, metode terbaiknya adalah SVM dengan nilai presisi mencapai 86,04%. Artinya, metode SVM adalah metode yang paling tepat untuk membuat prediksi kejadian La Niña dari seluruh nilai prediksi La Niña dalam model.

Sementara itu, metode dengan nilai *recall* terbaik adalah Naive-bayes. Nilai *recall* nya mencapai 88,63%. *Recall* sebesar 88,63,04% menunjukkan bahwa model berhasil mengidentifikasi 88,63% dari semua kejadian La Niña yang sebenarnya. Artinya, 11,37% dari fenomena La Niña tidak terdeteksi oleh model. Pada penelitian ini, tidak ada kondisi netral yang dipetakan. Oleh karena itu, ukuran *recall* memiliki peran penting untuk mengidentifikasi suatu kondisi tepat menjadi La Niña atau El Niño. Kemudian, Metode SVM juga memiliki nilai *F-1 score* tertinggi, yaitu sebesar 0,85. Ukuran *F-1 score* digunakan untuk mendeteksi metode yang memiliki performa baik dalam mendeteksi kelas positif dengan memperhatikan keseimbangan antara presisi dan *recall*. Apabila lewat parameter evaluasi sebelumnya, cenderung menghitung proporsi kejadian la nina maka ukuran *specificity* dapat menjadi gambaran keberhasilan klasifikasi El Niño.

Tabel 8. Perbandingan Performa Metode

| Data | Metode | Sensitivity | Specivity | Kappa |
|-------|-------------|-------------|-----------|-------|
| Train | KNN | 75,19 | 73,00 | 0,48 |
| | SVM | 80,45 | 79,50 | 0,60 |
| | Naïve Bayes | 81,95 | 70,25 | 0,52 |
| Test | KNN | 65,91 | 70,45 | 0,36 |
| | SVM | 84,09 | 86,36 | 0,70 |
| | Naïve Bayes | 88,63 | 70,68 | 0,48 |

Berdasarkan Tabel 8., metode SVM memiliki nilai *specificity* terbaik, yaitu 86,36%. Spesifisitas sebesar 86,36% menunjukkan bahwa dengan metode SVM dapat dengan benar mengidentifikasi 86,36% dari semua kasus yang sebenarnya El Niño. Nilai ini tidak jauh berbeda dengan nilai presisi dan *recall* yang didapat. Kemudian, nilai Cohen’s Kappa yang didapat menunjukkan bahwa model memiliki akurasi yang cukup baik. Secara konsisten nilai Kappa metode SVM lebih baik pada prediksi



data *training* maupun data *testing*. Nilai Kappa mengoreksi parametrik yang lebih sederhana seperti akurasi. Apabila dibandingkan, metode SVM secara konsisten memiliki nilai Kappa dan akurasi yang tinggi dibandingkan metode lainnya.

Hasil ini sejalan dengan penelitian [27], metode SVM diintegrasikan dengan ISOMAP untuk memprediksi fenomena El Niño dan La Niña. Kemudian, metode SVM juga dinilai memiliki komputasi yang efisien dalam mendeteksi kondisi kekeringan dan kebanjiran, dimana keduanya adalah dampak dari kondisi El Niño dan La Niña [28]. Namun, hasil pada penelitian ini tidak sejalan dengan penelitian yang dilakukan oleh [7]. Pada penelitian tersebut metode terbaik untuk klasifikasi fenomena El Niño dan La Niña adalah random forest. Sementara itu, pada penelitian ini tidak diujicobakan dengan metode random forest. Namun, performa klasifikasi metode SVM pada penelitian ini memiliki hasil yang cukup baik seperti penelitian sebelumnya.

Penggunaan rasio perbandingan 90% data *training* memungkinkan terjadinya *oversampling*. Berdasarkan Tabel 8., akurasi prediksi data *training* pada metode KNN lebih besar dari akurasi prediksi data *testing*, tetapi nilainya tidak melebihi 10%. Kemudian, akurasi prediksi data *testing* pada metode SVM dan Naive-bayes lebih besar dari akurasi prediksi data *training*. Artinya, metode tersebut terbebas dari indikasi *overfitting*.

IV. KESIMPULAN

Berdasarkan analisis data dan diskusi yang telah dilakukan, penelitian ini menggunakan metode KNN, SVM, dan Naive Bayes untuk mengklasifikasikan data terkait fenomena El Niño dan La Niña. Dalam penelitian ini, ditemukan bahwa rasio terbaik untuk klasifikasi El Niño dan La Niña di Indonesia adalah 90% data *training* dan 10% data *testing*. Hasil penerapan teknik data *mining* untuk klasifikasi fenomena El Niño dan La Niña di Indonesia menunjukkan bahwa metode SVM memiliki kinerja yang lebih baik dibandingkan dengan metode KNN dan Naive Bayes, baik dari segi nilai akurasi, presisi, maupun *F1-score*. Metode SVM memiliki nilai akurasi mencapai 85,23%, nilai presisi sebesar 86,04%, nilai *recall* sebesar 84,09%, dan 0,85 *F-1 score*. Oleh karena itu, metode SVM perlu diadopsi dalam sistem pemantauan iklim nasional untuk meningkatkan akurasi prediksi fase El Niño dan La Niña. Hal ini dapat membantu dalam mengidentifikasi perubahan iklim secara lebih akurat. Kemudian, saran bagi peneliti selanjutnya adalah perlunya proses eksplorasi lebih lanjut dalam pemilihan variabel independen serta dapat menggunakan data harian kiranya dapat meningkatkan akurasi klasifikasi.

UCAPAN TERIMA KASIH

Kami ingin mengucapkan terima kasih kepada semua pihak yang berperan dalam mendukung dan partisipasinya dalam penulisan artikel ini. Terimakasih kepada teman – teman SK2 yang telah memberikan kritik dan saran dalam kepenulisan artikel serta membantu dalam proses belajar pengambilan data.

REFERENSI

1. G. Technica, “Integrated Surveying for the Archaeological,” vol. 11, no. 2, pp. 39–50, 2016, doi: 10.21163/GT.
2. M. A. Rizaty, “10 Produsen Beras Terbesar Dunia, Indonesia Nomor 4,” 2022. <https://databoks.katadata.co.id/datapublish/2022/02/22/10-produsen-beras-terbesar-dunia-indonesia-nomor-4>.
3. N. Ramadhan, “ARTICLE REVIEW: The Influence Of Climate Change On Rice Production And Cultivation Patterns In Indonesia,” *J. Agron. Tanam. Trop.*, vol. 6, no. 1, 2024, doi: 10.36378/juatika.v6i1.3374.



4. BMKG, “Informasi ENSO.” <https://iklim.bmkg.go.id/id/enso/>.
5. N. Hayah, O. Soesanto, and M. A. Rahman, “Analysis of rainfall classification over Tanah Laut district based on global climate indicators using support vector machine method,” *J. Phys. Conf. Ser.*, vol. 2106, no. 1, 2021, doi: 10.1088/1742-6596/2106/1/012009.
6. N. Maher, T. P. Tabarin, and S. Milinski, “Combining machine learning and SMILEs to classify, better understand, and project changes in ENSO events,” *Earth Syst. Dyn.*, vol. 13, no. 3, pp. 1289–1304, 2022, doi: 10.5194/esd-13-1289-2022.
7. M. A. De Castro Santos, D. A. Vega-Oliveros, L. Zhao, and L. Berton, “Classifying El niño-southern oscillation combining network science and machine learning,” *IEEE Access*, vol. 8, pp. 55711–55723, 2020, doi: 10.1109/ACCESS.2020.2982035.
8. Q. Bi, K. E. Goodman, J. Kaminsky, and J. Lessler, “What is machine learning? A primer for the epidemiologist,” *Am. J. Epidemiol.*, vol. 188, no. 12, pp. 2222–2239, 2019, doi: 10.1093/aje/kwz189.
9. G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “KNN model-based approach in classification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2888, no. January, pp. 986–996, 2003, doi: 10.1007/978-3-540-39964-3_62.
10. V. Bijalwan, V. Kumar, P. Kumari, and J. Pascual, “KNN based machine learning approach for text and document mining,” *Int. J. Database Theory Appl.*, vol. 7, no. 1, pp. 61–70, 2014, doi: 10.14257/ijtda.2014.7.1.06.
11. W. Xing and Y. Bei, “Medical Health Big Data Classification Based on KNN Classification Algorithm,” *IEEE Access*, vol. 8, pp. 28808–28819, 2020, doi: 10.1109/ACCESS.2019.2955754.
12. K. Mohammadi and N. Goudarzi, “Study of inter-correlations of solar radiation, wind speed and precipitation under the influence of El Niño Southern Oscillation (ENSO) in California,” *Renew. Energy*, vol. 120, pp. 190–200, 2018, doi: 10.1016/j.renene.2017.12.069.
13. N. J. Kuo and C. R. Ho, “ENSO effect on the sea surface wind and sea surface temperature in the Taiwan Strait,” *Geophys. Res. Lett.*, vol. 31, no. 13, pp. 10–13, 2004, doi: 10.1029/2004GL020303.
14. C. hoon Hong, K. D. Cho, and H. J. Kim, “The relationship between ENSO events and sea surface temperature in the east (Japan) sea,” *Prog. Oceanogr.*, vol. 49, no. 1–4, pp. 21–40, 2001, doi: 10.1016/S0079-6611(01)00014-3.
15. M. L. L’Heureux, D. C. Collins, and Z. Z. Hu, “Linear trends in sea surface temperature of the tropical Pacific Ocean and implications for the El Niño-Southern Oscillation,” *Clim. Dyn.*, vol. 40, no. 5–6, pp. 1223–1236, 2013, doi: 10.1007/s00382-012-1331-2.
16. S. Curtis and R. Adler, “ENSO indices based on patterns of satellite-derived precipitation,” *J. Clim.*, vol. 13, no. 15, pp. 2786–2793, 2000, doi: 10.1175/1520-0442(2000)013<2786:EIBOPO>2.0.CO;2.
17. W. Su, N. G. Loeb, L. Liang, N. Liu, and C. Liu, “The El Niño-Southern Oscillation effect on tropical outgoing longwave radiation: A daytime versus nighttime perspective,” *J. Geophys. Res.*, vol. 122, no. 15, pp. 7820–7833, 2017, doi: 10.1002/2017JD027002.
18. K. G. Pavlakis *et al.*, “ENSO surface longwave radiation forcing over the tropical Pacific,” *Atmos. Chem. Phys.*, vol. 7, no. 8, pp. 2013–2026, 2007, doi: 10.5194/acp-7-2013-2007.
19. R. S. Nerem, D. P. Chambers, C. Choe, and G. T. Mitchum, “Estimating Mean Sea Level Change from the TOPEX and Jason Altimeter Missions,” *Mar. Geod.*, vol. 33, pp. 435–446, 2010, doi: 10.1080/01490419.2010.491031.
20. L. S. Supriatin and M. Martono, “Impacts of Climate Change (El Nino, La Nina, and Sea Level) on the Coastal Area of Cilacap Regency,” *Forum Geogr.*, vol. 30, no. 2, p. 106, 2016, doi: 10.23917/forgeo.v30i2.2449.
21. A. Yananto and R. M. Sibarani, “ANALISIS KEJADIAN EL NINO DAN PENGARUHNYA TERHADAP INTENSITAS CURAH HUJAN DI WILAYAH JABODETABEK (Studi Kasus : Periode Puncak Musim Hujan Tahun 2015/2016),” *J. Sains Teknol. Modif. Cuaca*, vol. 17, no. 2, p. 65, 2016, doi: 10.29122/jstmc.v17i2.541.
22. T. Zhang, W. Zhang, F. Jiang, and F. F. Jin, “La Niña’s Teleconnection to the Indian Ocean Dipole Controlled by Its Longitudinal Position,” *Geophys. Res. Lett.*, vol. 51, no. 3, pp. 1–10, 2024, doi: 10.1029/2023GL107119.
23. BMKG, *PERATURAN BADAN METEOROLOGI, KLIMATOLOGI, DAN GEOFISIKA REPUBLIK INDONESIA NOMOR 9 TAHUN 2019 TENTANG PENYEDIAAN DAN PENYEBARAN PERINGATAN DINI IKLIM EKSTRIM.* .
24. J. S. Kug, F. F. Jin, and S. Il An, “Two types of El Niño events: Cold tongue El Niño and warm pool El Niño,” *J. Clim.*, vol. 22, no. 6, pp. 1499–1515, 2009, doi: 10.1175/2008JCLI2624.1.



SENADA
Seminar Nasional Sains Data

Seminar Nasional Sains Data 2024 (SENADA 2024)

E-ISSN 2808-5841

UPN “Veteran” Jawa Timur

P-ISSN 2808-7283

25. G. Guimarães Nobre, S. Muis, T. I. E. Veldkamp, and P. J. Ward, “Achieving the reduction of disaster risk by better predicting impacts of El Niño and La Niña,” *Prog. Disaster Sci.*, vol. 2, p. 100022, 2019, doi: 10.1016/j.pdisas.2019.100022.
26. Badan Meteorologi Klimatologi dan Geofisika (BMKG), “Perkiraan Musim Hujan 2023/2023,” *Angew. Chemie Int. Ed.* 6(11), 951–952., no. September, pp. 1–94, 2023.
27. J. C. H. Tseng, “An ISOMAP Analysis of Sea Surface Temperature for the Classification and Detection of El Niño & La Niña Events,” *Atmosphere (Basel)*., vol. 13, no. 6, 2022, doi: 10.3390/atmos13060919.
28. G. Yin, T. Yoshikane, R. Kaneko, and K. Yoshimura, “Improving Global Subseasonal to Seasonal Precipitation Forecasts Using a Support Vector Machine-Based Method *Journal of Geophysical Research : Atmospheres*,” pp. 1–17, 2023, doi: 10.1029/2023JD038929.