



## Perbandingan Metode Klasifikasi Untuk Mengestimasi *Land Surface Temperature* di Jawa Barat

Alifian Wahyu Prakhoso<sup>1</sup>, Nailu Rokhmah<sup>2</sup>, Reni Pratamawati<sup>3</sup>, Robert Kurniawan<sup>4</sup>

<sup>1, 2, 3</sup>Program Studi D-IV Statistik, Politeknik Statistika STIS  
<sup>4</sup>Program Studi D-IV Komputasi Statistik, Politeknik Statistika STIS

<sup>1</sup>[212111876@stis.ac.id](mailto:212111876@stis.ac.id)

<sup>2</sup>[212112240@stis.ac.id](mailto:212112240@stis.ac.id)

<sup>3</sup>[212112313@stis.ac.id](mailto:212112313@stis.ac.id)

<sup>4</sup>[robertk@stis.ac.id](mailto:robertk@stis.ac.id)

Corresponding author email:[212112240@stis.ac.id](mailto:212112240@stis.ac.id)

**Abstract:** In 2023 is set to be the hottest year globally compared to pre-industrial levels. It is predicted that in the first half of 2030, global warming will reach 1.5°C above the pre-industrial period. The West Java region in October 2023 had the highest monthly average temperature at the Penggung Meteorological Post, Cirebon and the Kertajati Meteorological Station, Majalengka. Apart from that, West Java channels greenhouse gas emissions through industrial areas in the area. This research aims to look at the description of land surface temperatures in West Java in 2023 and identify the best method for classifying land surface temperatures in West Java in 2023. Data obtained from Landsat-8 and Sentinel-2 satellite images in the form of NDVI, NDBI and NDWI. The method used is supervised learning in the form of classification, including Random Forest, Decision Tree, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Naïve Bayes. The conclusion of this research shows that the Support Vector Machine (SVM) method provides excellent results with the highest accuracy value of 93.70 percent and the highest kappa value of 87.36 percent. Therefore, the Support Vector Machine (SVM) method can be used as a reference for classification.

**Keywords:** Land Surface Temperature, Comparison, Classification, Data Mining, SVM

**Abstrak:** Tahun 2023 ditetapkan sebagai tahun terpanas global dibandingkan dengan pra industri. Diprediksi pada paruh pertama tahun 2030, pemanasan global mencapai 1,5°C diatas periode pra industri. Wilayah Jawa Barat pada bulan Oktober 2023 memiliki suhu rata-rata bulan tertinggi di Pos Meteorologi Penggung, Cirebon dan Stasiun Meteorologi Kertajati, Majalengka. Selain itu Jawa Barat menyumbang emisi gas rumah kaca melalui kawasan-kawasan industri di daerah tersebut. Penelitian ini bertujuan untuk melihat gambaran suhu permukaan tanah di Jawa Barat tahun 2023 dan mengidentifikasi metode klasifikasi suhu permukaan tanah di Jawa Barat tahun 2023 yang terbaik. Data diperoleh dari citra satelit Landsat-8 dan Sentinel-2 berupa NDVI, NDBI, dan NDWI. Metode yang digunakan adalah *supervised learning* berupa klasifikasi antara lain, *Random Forest*, *Decision Tree*, *K-Nearest Neighbor (KNN)*, *Support Vector Machine (SVM)*, dan *Naïve Bayes*. Kesimpulan dari penelitian ini menunjukkan bahwa metode *Support Vector Machine (SVM)* memberikan hasil sangat baik dengan nilai akurasi tertinggi sebesar 93,70 persen dan nilai kappa tertinggi sebesar 87,36 persen. Oleh karena itu, metode *Support Vector Machine (SVM)* dapat digunakan sebagai acuan untuk melakukan klasifikasi.

**Kata kunci:** Suhu Permukaan Tanah, Perbandingan, Klasifikasi, Data Mining, SVM

### I. PENDAHULUAN

Tahun 2023, secara global ditetapkan sebagai tahun terpanas dibandingkan dengan pengamatan 174 tahun yang lalu. Tercatat hingga bulan Oktober 2023 suhu global mencapai 1,48°C lebih hangat daripada rata-rata suhu periode pra industri tahun 1850 hingga 1990 dengan anomali sebesar 0,12°C. Sebelumnya tahun terpanas terjadi pada tahun 2016 dengan suhu rata rata sebesar 1,29°C. Selain itu pernah juga terjadi pada tahun 2020 dengan suhu rata-rata mencapai 1,27°C. Kemudian pada 10 tahun terakhir periode 2014-2023, rata-rata suhu global mencapai 1,19°C diatas rata-rata periode 1850-1900.

Diperkirakan pada paruh pertama tahun 2030, pemanasan global terjadi mencapai 1,5°C diatas periode pra industri (WMO, 2023) [1]. Namun, apabila emisi gas rumah kaca dalam skenario sangat tinggi dan kadar CO<sub>2</sub> naik memungkinkan global mencapai 2°C lebih hangat dari pra industri yang terjadi pada akhir tahun 2020-an. IPCC [2] menyatakan bahwa skenario emisi gas rumah kaca yang rendah dan sangat rendah dapat menurunkan emisi CO<sub>2</sub> nol bersih pada tahun 2050 dan 2070. Kawasan Asia Tenggara dan Pasifik menjadi kawasan dengan penyumbang 8% emisi CO<sub>2</sub> dari tahun 1850-2019.

Suhu udara permukaan rata-rata bulanan di Indonesia dari Januari hingga Desember diperkirakan sebesar 0,3°C lebih hangat dibanding periode 1991-2020 [3]. Dengan anomali antara +0,23°C hingga



+0,36°C. Peningkatan suhu diperkirakan akibat beberapa daerah mengalami hujan tahunan dibawah angka normal. Hal ini menyebabkan daerah mengalami kekeringan dan minimnya tutupan awan pada siang hari sehingga berpotensi mengalami suhu permukaan tanah yang lebih tinggi. Peningkatan suhu rata-rata nasional seiring dengan meningkatnya emisi gas rumah kaca dan CO<sub>2</sub> dari tahun ke tahun.

BMKG merilis data anomali suhu udara rata-rata per-stasiun pada bulan Oktober 2023 menunjukkan anomali positif. Kemudian terdapat nilai suhu rata-rata bulan Oktober 2023 lima tertinggi yaitu di Stasiun Klimatologi Jawa Tengah (Semarang), Pos Meteorologi Penggung (Cirebon), Stasiun Meteorologi Kertajati (Majalengka), Stasiun Meteorologi Perak I (Surabaya), dan Stasiun Meteorologi Maritim Tanjung Emas. Dari data tersebut terdapat dua stasiun meteorologi yang terletak di Jawa Barat. Diperkirakan wilayah ini mengalami peningkatan suhu rata-rata bulanan maupun tahunan.

Jawa Barat juga memiliki 42 kawasan industri yang telah memiliki Izin Usaha Kawasan Industri (IUKI) dan sedang beroperasi. Hal ini dapat berdampak pada pemanasan lokal dan peningkatan suhu rata-rata nasional. Hal tersebut terjadi seiring dengan aktivitas industri yang berkontribusi menyebabkan emisi gas rumah kaca karena pembakaran bahan bakar untuk alat industri. Selain itu perubahan penggunaan lahan menjadi kawasan industri akan meningkatkan karbon CO<sub>2</sub>.

Jawa Barat secara administratif terbagi dalam 627 wilayah kecamatan. Sedangkan luas wilayah Jawa Barat adalah berupa daratan seluas 35.377,76 km<sup>2</sup>. Wilayah ini berada pada 5°50`-7°50` Lintang Selatan dan 104°48`-108°48` Bujur Timur. Wilayah Provinsi Jawa Barat bagian utara berbatasan dengan Laut Jawa, bagian selatan berbatasan dengan Samudera Hindia, bagian barat berbatasan dengan Provinsi Banten dan Provinsi DKI Jakarta, dan bagian timur berbatasan dengan Provinsi Jawa Tengah. Wilayah ini merupakan daratan yang dibedakan atas wilayah pegunungan curam di selatan dengan ketinggian lebih dari 1.500 mdpl, wilayah lereng bukit yang landai di tengah dengan ketinggian 100-1.500 mdpl, wilayah dataran luas di utara dengan ketinggian 0-10 mdpl, dan wilayah aliran sungai. Tahun 2023 BPS mencatat bahwa jumlah penduduk di Jawa Barat mencapai 49,86 juta jiwa. Maka dapat diketahui kepadatan penduduk sebanyak 1.346,12 jiwa/km<sup>2</sup> [4].

Provinsi Jawa Barat pada tahun 2023 mencatat penduduk sebanyak 49,86 juta jiwa. Selain itu Jawa Barat memiliki laju pertumbuhan penduduk sebesar 1,32% terhadap tahun 2022. Hal ini beriringan dengan luas lahan hijau yang digunakan untuk menunjang kehidupan masyarakat seperti membangun hunian, fasilitas jalan raya, pabrik industri, hingga pasar. Perubahan fungsi lahan ini dapat mengurangi vegetasi daerah tersebut yang berdampak kurangnya tempat resapan air, kekuatan tanah berkurang, daratan terkikis sehingga terjadi bencana seperti banjir, kekeringan, dan kenaikan permukaan air laut.

Penelitian Wan Li dkk (2019) [6] menunjukkan metode ANN dan RF mempunyai akurasi yang tinggi pada estimasi suhu permukaan tanah di Beijing. Selain itu, penelitian Shafizadeh dkk (2020) [7] menyatakan metode klasifikasi *Random Forest*, *Generalized Additive Model (GAM)*, *Boosted Regression Tree (BRT)*, dan *Support Vector Machine* menunjukkan R Square sebesar 60,6-62,1%. Keempat metode tersebut tidak signifikan mengalami perbedaan. Kemudian penelitian dari Cantika dkk (2020) [8] menunjukkan performa metode klasifikasi *K-NN*, *CART*, *RF*, *SVM*, dan *LDA* memiliki nilai *Kappa* sebesar 0,8 hingga 1 untuk mengidentifikasi risiko kekeringan. Namun, penelitian mengenai komparasi metode klasifikasi suhu permukaan tanah di Indonesia masih terbatas, terutama di lokus Jawa Barat. Oleh karena itu, penelitian ini menggunakan *machine learning* jenis klasifikasi yaitu *Random Forest*, *Decision Tree*, *K-Nearest Neighbor (KNN)*, *Support Vector Machine (SVM)*, dan *Naïve Bayes* untuk mengidentifikasi pola, kemudian memprediksi hasil dengan data baru, dan membandingkan hasil evaluasi dari lima model yang terbentuk.



Penelitian ini bertujuan untuk melihat gambaran suhu permukaan tanah di Jawa Barat tahun 2023 dan mengidentifikasi metode klasifikasi suhu permukaan tanah antara lain, *Random Forest*, *Decision Tree*, *K-Nearest Neighbour (KNN)*, *Support Vector Machine (SVM)*, dan *Naïve Bayes* di Jawa Barat tahun 2023 yang terbaik. Penelitian ini juga berkontribusi untuk penanganan perubahan iklim sesuai dengan tujuan ke-13 Tujuan Pembangunan Berkelanjutan. Kemudian ikut berkomitmen membatasi kenaikan suhu rata-rata global kurang dari 2°C diatas tingkat pra industri dan menjaga suhu rata-rata global saat ini sebesar 1,5°C diatas tingkat pra industri agar tidak mengalami kenaikan pada periode kedepan. Kedua kebijakan tersebut melibatkan semua negara untuk berkontribusi menurunkan emisi gas rumah kaca dan adaptasi terhadap perubahan iklim. Selain itu, penelitian ini ikut mendukung sasaran kelima pembangunan RPJPN 2025-2045 atau dikenal dengan Indonesia Emas yaitu menurunkan intensitas emisi gas rumah kaca menuju net zero emission dengan target penurunan intensitas emisi gas rumah kaca sebesar 93,5%.

## II. LANDASAN TEORI DAN METODE

Proses pengumpulan data dilakukan dengan menggunakan *Google Earth Engine (GEE)*. Peta *shapefile* batas kecamatan Provinsi Jawa Barat diinputkan untuk digunakan sebagai peta dasar pemrosesan data. Kemudian Satelit Landsat-8 digunakan untuk menangkap data LST dan Satelit Sentinel-2 digunakan untuk menangkap data NDVI, NDBI dan NDWI. Kemudian, masing-masing data didapatkan dengan menggunakan *syntax* yang sesuai. Data yang didapatkan melalui GEE tersebut lalu di-*export* ke dalam format *Geotiff* yang kemudian diekstrak menggunakan *software* Geoda. Dengan memanfaatkan peta *shapefile* batas kecamatan Provinsi Jawa Barat dan fitur *Zonal Statistics* pada Geoda, dilakukan proses penggabungan data sehingga didapatkan data-data tersebut dalam satu file berbentuk *excel*. Setelah itu, dilakukan analisis data dengan metode klasifikasi menggunakan *Software R Studio*.

### II.1. Land Surface Temperature (LST)

LST adalah suhu efektif dari permukaan tanah yang memancarkan radiasi termal yang dapat diukur dari sensor satelit [9]. Dalam penerapannya, LST sering digunakan untuk menganalisis perubahan lingkungan khususnya suhu dan iklim. Rumus dari LST adalah sebagai berikut.

$$LST = \frac{T}{1 + \left(\frac{\lambda T}{\rho}\right) \ln(\epsilon)} \quad (1)$$

Menurut pengukuran suhu, daerah tropis mempunyai suhu rata-rata 20 derajat. Sementara itu Indonesia sebagai negara yang berada di daerah tropis yang umumnya mempunyai rata-rata suhu mencapai 35 derajat dengan tingkat kelembaban tinggi mencapai 85 persen yang dapat disebut iklim tropis panas lembab [10]. Oleh karena itu, *Land Surface Temperature (LST)* dapat diklasifikasikan menjadi suhu normal untuk wilayah yang mempunyai suhu mencapai 35°C sedangkan wilayah yang mempunyai suhu mencapai lebih dari 35°C dianggap sebagai wilayah yang mempunyai suhu tidak normal seperti pada tabel 1 berikut.

Tabel 1. Klasifikasi Nilai LST

Kelas	LST	Keterangan
1	≤ 35°C	Suhu Normal
2	> 35°C	Suhu Tidak Normal

### II.2. Normalized Difference Vegetation Index (NDVI)

Menurut Rouse, dkk [11] NDVI merupakan indeks yang digunakan untuk mengukur kerapatan vegetasi dengan metode penginderaan jauh. Indeks ini dihitung dengan menggunakan perbedaan antara



cahaya inframerah dekat (NIR) yang dipantulkan vegetasi dan cahaya merah (RED) yang diserap vegetasi. Rumus dari NDVI sebagai berikut.

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (2)$$

Nilai NDVI berkisar antara -1,0 hingga +1,0. Nilai positif menunjukkan keberadaan vegetasi yang lebih padat. sedangkan, nilai negatif menunjukkan keberadaan air atau tanah terbuka. Selain itu, nilai yang mendekati nol (coklat) menunjukkan area tanah gundul, dan nilai negatif (biru) menunjukkan keberadaan air.

### II.3. Normalized Difference Building Index (NDBI)

NDBI adalah indeks yang digunakan untuk mengidentifikasi daerah terbangun (urban) dengan menggunakan perbedaan antara pantulan inframerah pendek (SWIR) dan pantulan inframerah dekat (NIR) [12]. Rumus dari NDBI adalah

$$NDBI = \frac{SWIR-NIR}{SWIR+NIR} \quad (3)$$

Nilai NDBI berkisar antara -1,0 hingga 1. Nilai NDBI yang lebih tinggi menunjukkan kepadatan bangunan yang tinggi, sedangkan nilai yang lebih rendah menunjukkan vegetasi, lahan pertanian atau perairan. Nilai NDBI dengan variasi antara -0,18 hingga 0,02 yang sangat dekat dengan 0 menunjukkan sel tersebut hanya mengandung air.

### II.4. Normalized Difference Water Index (NDWI)

NDWI adalah indeks yang digunakan untuk mengidentifikasi air terbuka melalui citra satelit. NDWI dihitung berdasarkan perbedaan antara cahaya hijau (GREEN) dan cahaya inframerah dekat (NIR) [13]. Rumus NDWI adalah sebagai berikut

$$NDWI = \frac{GREEN-NIR}{GREEN+NIR} \quad (4)$$

Nilai NDWI berkisar antara -1 hingga +1. Nilai keberadaan air atau tanah terbuka berada pada angka negatif.

### II.5. Random Forest

Menurut Breiman [14] *Random Forest* merupakan salah satu metode klasifikasi yang terdiri dari kumpulan pohon keputusan yang dihasilkan secara acak. Setiap pohon keputusan tersebut menghasilkan prediksi kelas dan kelas yang paling banyak dipilih oleh pohon keputusan akan menjadi prediksi akhir dari model. Keunggulan dari metode klasifikasi ini adalah risiko overfitting yang lebih kecil jika dibandingkan dengan *decision tree*. Hal tersebut karena keputusan pada setiap tingkatan pohon saling terikat. Kekurangan dari metode ini adalah kecepatan dalam mengolah data berskala besar yang cenderung lebih lambat. Padahal metode data mining diterapkan pada data yang *on time*. Metode *random forest* dipilih sebagai salah satu metode klasifikasi yang digunakan karena menurut Xu et al [25] metode *random forest* memiliki keunggulan mampu meningkatkan akurasi LST dibandingkan metode lain.

### II.6. Decision Tree

*Decision Tree* adalah metode *machine learning* yang digunakan untuk mengklasifikasikan dengan memetakan seluruh dataset (*root node*) ke dalam node-node yang dibagi berdasarkan kriteria tertentu (*decision nodes*). Node terakhir akan menjadi keputusan atau hasil klasifikasi (*leaf nodes*) [15]. *Decision Tree* memiliki beberapa keunggulan dibandingkan metode yang lain, seperti daerah pengambilan keputusan yang lebih sederhana dan spesifik. Kemudian sampel diuji hanya berdasarkan



kriteria tertentu sehingga perhitungan yang tidak diperlukan di eliminasi. Namun, karena daerah pengambilan keputusan sederhana sehingga jumlah error pada setiap node atau cabang cukup besar. Oleh karena itu, terdapat kesulitan dalam menemukan model *decision tree* yang paling optimal. Menurut penelitian Kim et al [26] *decision tree* efektif dalam melakukan klasifikasi status beku/cair pada permukaan tanah berdasarkan karakteristik emisi dan hamburan sehingga mampu meningkatkan akurasi pengukuran LST. Oleh karena itu, metode *decision tree* dipilih sebagai salah satu metode klasifikasi yang digunakan.

### II.7. K-Nearest Neighbor

Menurut Altman [16] *K-Nearest Neighbor* adalah metode non-parametrik yang dapat digunakan untuk klasifikasi dengan menentukan banyaknya jumlah tetangga terdekat. Metode ini memiliki asumsi bahwa titik data yang serupa berada di dekat satu dengan yang lain. Setelah menentukan jumlah tetangga, dilakukan perhitungan jarak, biasa yang digunakan yaitu jarak *euclidean*, kemudian data yang berada dalam jangkauan jarak tersebut akan dikelompokkan sebagai tetangganya. KNN memiliki keunggulan yaitu cocok digunakan untuk data yang berjumlah besar sehingga klasifikasi yang dilakukan akan lebih akurat. Namun, terdapat kesulitan dalam menentukan  $k$  atau jumlah tetangga yang dapat menghasilkan model yang optimal. Pada penelitian ini, metode KNN digunakan karena memiliki keunggulan ketika diterapkan dalam data penginderaan jauh sehingga memberikan klasifikasi dengan akurasi yang tinggi [27].

### II.8. Support Vector Machine (SVM)

*Support Vector Machine* merupakan salah satu teknik *machine learning* klasik yang dapat membantu menyelesaikan masalah klasifikasi big data terutama dapat membantu aplikasi multi domain dalam lingkungan big data [17]. Prinsip dasar dari SVM adalah *linear classifier* yang selanjutnya dikembangkan agar dapat mencakup permasalahan *non-linear* dengan konsep *kernel trick* pada ruang berdimensi tinggi. Tujuan dari metode ini adalah menemukan *hyperplane* atau fungsi pemisah linear yang digunakan untuk memisahkan dua kategori data dalam ruang  $n$ -dimensi. Oleh karena itu, SVM memiliki keunggulan dalam mengklasifikasi dalam ruang dimensi tinggi dengan akurasi yang tinggi dan akurat. Namun, SVM juga memiliki kekurangan yaitu kurang mampu mengklasifikasi kelas yang cenderung tumpang tindih dan memerlukan proses yang panjang. Menurut Tnanh dan Kappas [27] metode SVM memberikan hasil klasifikasi dengan akurasi yang tinggi dalam penginderaan jauh karena mampu memaksimalkan batas-batas antar kelas. Oleh karena itu, metode ini dipilih sebagai salah satu metode klasifikasi yang digunakan.

### II.9. Naïve Bayes

*Naïve Bayes* merupakan metode klasifikasi yang menggunakan pengklasifikasian probabilitas berdasarkan teori keputusan Bayesian dengan asumsi independensi yang naive [18]. Metode klasifikasi ini memiliki keunggulan selain sederhana juga cocok dalam kasus yang terdapat probabilitas bersyarat. Namun, hal tersebut juga menjadikan kelemahan dimana memungkinkan terjadi nilai probabilitas nol dalam data training. Permasalahan tersebut dapat ditangani dengan melakukan *smoothing*. Kemudian asumsi independensi mengakibatkan akurasi dari prediksi berkurang karena umumnya terdapat korelasi antar atribut atau variabel. Menurut penelitian Shafi et al [28] metode *naïve bayes* memberikan hasil klasifikasi yang baik dan akurat ketika digunakan dalam memprediksi efek *Urban Heat Island* (UHI).

Tahapan pengujian metode diawali dengan melakukan pembagian dataset menjadi data training dan data testing. Pembagian dataset tersebut dilakukan dengan perbandingan 80:20 yang mana 80 persen dari dataset merupakan data training serta 20 persen dari dataset merupakan data testing. Perbandingan



pembagian dataset tersebut merupakan perbandingan yang paling umum digunakan [19]. Setelah itu, dilakukan tahapan evaluasi model menggunakan overall accuracy dan nilai kappa dari model klasifikasi *Random Forest*, *Decision Tree*, *K-Nearest Neighbour*, *Support Vector Machine*, dan *Naïve Bayes*. Evaluasi model ini digunakan untuk mengukur seberapa baik kinerja model klasifikasi dengan menggunakan *confusion matrix*. *Confusion matrix* ini berisi mengenai perbandingan hasil klasifikasi prediksi dengan hasil klasifikasi aktual. Kemudian dari *confusion matrix* dapat dihitung nilai akurasi dan nilai kappanya. *Confusion matrix* digunakan sebagai pengelompokan hasil prediksi model menjadi beberapa kategori berdasarkan nilai prediksi dan aktualnya seperti yang ditunjukkan pada tabel 2.

**Tabel 2.** *Confusion Matrix*

		Aktual	
		Normal	Tidak Normal
Prediksi	Normal	True Positif	False Positif
	Tidak Normal	False Negatif	True Negatif

Akurasi merupakan persentase prediksi benar dari semua prediksi. Berikut ini rumus untuk menghitung nilai akurasi.

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (5)$$

Keterangan:

TP : True Positive

FP : False Positive

FN : False Negative

TN : True Negative

*True Positive* (TP) merupakan kasus untuk model yang memprediksi positif dan hasil aktualnya positif, *False Positive* (FP) merupakan kasus untuk model yang memprediksi positif tetapi hasil aktualnya negatif, *False Negative* (FN) merupakan kasus untuk model yang memprediksi positif tetapi hasil aktualnya negatif, dan *True Negative* (TN) merupakan kasus untuk model yang memprediksi negatif dan hasil aktualnya negatif. Secara umum, nilai akurasi diinterpretasikan seperti pada tabel 3.

**Tabel 3.** Interpretasi Nilai Akurasi

Nilai Accuracy	Interpretasi
>90%	Sangat Baik
80-90%	Baik
70-80%	Cukup Baik
<70%	Buruk

Kappa atau Cohen’s kappa merupakan ukuran kesepakatan dinormalisasi yang mempertimbangkan kesempatan kebetulan dalam data. Berikut ini rumus untuk menghitung nilai kappa.

$$Kappa = \frac{Pa-Pc}{1-Pc} \quad (6)$$

Keterangan:

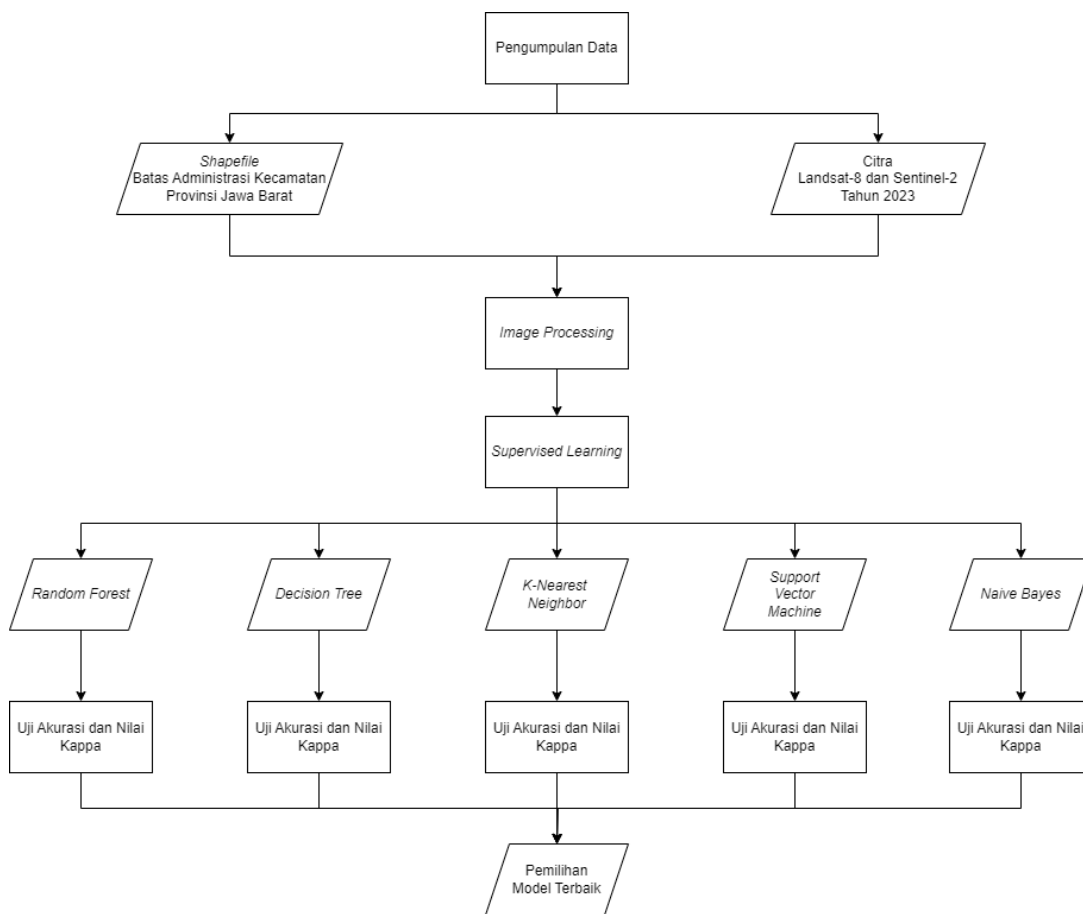
Pa : Proporsi kesepakatan teramati

Pc : Proporsi kesepakatan harapan

Selain itu, skala untuk melakukan pengukuran nilai kappa seperti yang dilakukan oleh beberapa peneliti seperti pada tabel 4 berikut.

**Tabel 4.** Skala Pengukuran Nilai Kappa Menurut Altman (1992)

Nilai Kappa	Interpretasi
<20%	Poor
21-40%	Fair
41-60%	Moderate
61-80%	Good
81-100%	Very Good



**Gambar 1.** Alur Pengolahan

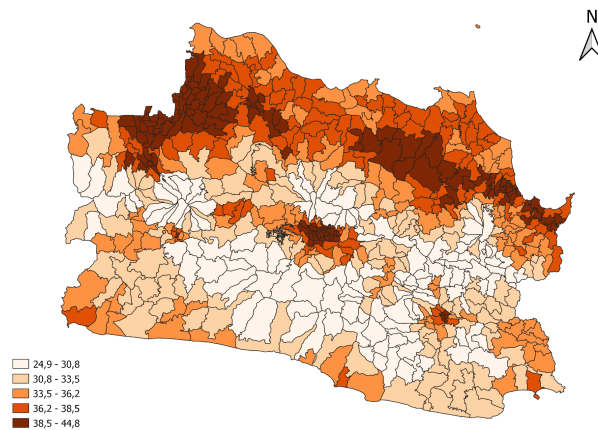
Berdasarkan gambar 1, dijelaskan mengenai alur pengolahan yang digunakan dalam penelitian ini. Proses pengumpulan data melalui GEE dengan memanfaatkan *shapefile* batas kecamatan Provinsi Jawa Barat serta Satelit Landsat-8 dan Satelit Sentinel-2. Data yang diperoleh dari GEE berformat *Geotiff* kemudian dilakukan tahap *image processing* untuk mendapat data berformat *excel* melalui *software* Geoda. Setelah itu, dilakukan pemodelan *supervised learning* dengan melakukan pengklasifikasian menggunakan *Random Forest*, *Decision Tree*, *K-Nearest Neighbour*, *Support Vector Machine*, dan *Naïve Bayes*. Pemilihan model terbaik didasarkan pada nilai akurasi dan nilai kappa yang paling besar.

### III. PEKERJAAN DAN DISKUSI HASIL

#### III.1. Hasil

Berdasarkan gambar 2, secara umum dapat terlihat bahwa suhu permukaan tanah (LST) Provinsi Jawa Barat tahun 2023 cukup beragam. Hal tersebut ditunjukkan dengan peta tematik dimana semakin gelap warna suatu wilayah maka semakin tinggi pula nilai suhu permukaan tanah di daerah tersebut.

Kemudian wilayah-wilayah tersebut cenderung berkelompok dengan wilayah yang memiliki suhu permukaan tanah yang tidak jauh berbeda. Selain itu, daerah dengan suhu permukaan tanah yang tinggi sebagian besar berada di daerah bagian utara Provinsi Jawa Barat serta daerah metropolitan seperti Bandung, Bekasi, Depok, dan Bogor. Hal ini karena daerah metropolitan terjadi fenomena pulau panas perkotaan (*Urban Heat Island*) sehingga suhu permukaannya menjadi lebih tinggi dibandingkan daerah sekitarnya.



**Gambar 2.** Peta Tematik Persebaran LST Provinsi Jawa Barat

**Tabel 5.** Confusion Matrix

Metode	Prediksi	Aktual	
		Normal	Tidak Normal
<i>Random Forest</i>	Normal	50	5
	Tidak Normal	5	67
<i>Decision Tree</i>	Normal	59	9
	Tidak Normal	4	55
<i>K-Nearest Neighbor</i>	Normal	60	7
	Tidak Normal	1	59
<i>Support Vector Machine</i>	Normal	58	5
	Tidak Normal	2	62
<i>Naïve Bayes</i>	Normal	58	11
	Tidak Normal	5	53

Berdasarkan tabel 5 dapat diketahui mengenai confusion matrix dari pengelompokan hasil prediksi model oleh algoritma *Random Forest*, *Decision Tree*, *K-Nearest Neighbour*, *Support Vector Machine*, dan *Naïve Bayes*. Model *Random Forest* mampu memprediksi suhu normal dan hasil aktualnya juga normal sebanyak 50 data. Selain itu, model ini mampu memprediksi suhu tidak normal dan hasil aktualnya juga tidak normal sebanyak 67 data. Sementara itu, model *Random Forest* dapat melakukan kesalahan prediksi yang mana model memprediksi suhu normal tetapi hasil aktualnya suhu tidak normal sebanyak 5 data. Selain itu, kesalahan lain dalam prediksi model ini yaitu prediksi suhu tidak normal tetapi hasil aktualnya suhu normal sebanyak 5 data.





Sementara itu, model *Decision Tree* yang didapatkan berhasil memprediksi suhu normal dan hasil aktualnya juga normal sebanyak 59 data. Selain itu, model ini mampu memprediksi suhu tidak normal ketika hasil aktualnya juga tidak normal sebanyak 55 data. Namun, model ini melakukan kesalahan prediksi di mana ia memprediksi suhu normal, tetapi hasil aktualnya adalah suhu tidak normal sebanyak 9 data. Kesalahan lainnya adalah ketika model memprediksi suhu tidak normal, tetapi hasil aktualnya adalah suhu normal sebanyak 4 data.

Selanjutnya, model *K-Nearest Neighbor* yang didapatkan mampu memberikan prediksi suhu normal dan hasil aktualnya juga normal sebanyak 60 data. Selain itu, model ini juga mampu memprediksi suhu tidak normal dan hasil aktualnya juga tidak normal sebanyak 59 data. Sementara itu, model *K-Nearest Neighbor* dapat melakukan kesalahan prediksi yang mana model memprediksi suhu normal tetapi hasil aktualnya suhu tidak normal sebanyak 7 data. Kemudian, kesalahan lain dalam prediksi model ini yaitu prediksi suhu tidak normal tetapi hasil aktualnya suhu normal sebanyak 1 data.

Kemudian, model *Support Vector Machine* yang didapatkan mampu memprediksi suhu normal yang sesuai dengan hasil aktualnya sebanyak 58 data. Selain itu, model ini juga mampu memprediksi suhu tidak normal yang sesuai dengan hasil aktualnya sebanyak 62 data. Sementara itu, model *Support Vector Machine* dapat melakukan kesalahan prediksi yang mana model memprediksi suhu normal tetapi hasil aktualnya suhu tidak normal sebanyak 5 data. Selanjutnya, kesalahan lain dalam prediksi model ini yaitu prediksi suhu tidak normal tetapi hasil aktualnya suhu normal sebanyak 2 data.

Selain itu, model *Naïve Bayes* yang didapatkan berhasil memprediksi suhu normal yang sesuai dengan aktualnya sebanyak 58 data. Selain itu, model ini juga mampu memprediksi suhu tidak normal dan hasil aktualnya juga tidak normal sebanyak 53 data. Sementara itu, model *Naïve Bayes* dapat melakukan kesalahan prediksi yang mana model memprediksi suhu normal tetapi hasil aktualnya suhu tidak normal sebanyak 11 data. Kemudian, kesalahan lain dalam prediksi model ini yaitu prediksi suhu tidak normal tetapi hasil aktualnya suhu normal sebanyak 5 data.

Setelah itu, dilakukan penghitungan nilai akurasi dan kappa menggunakan *confusion matrix*. Nilai akurasi digunakan untuk melihat gambaran umum seberapa baik model mengklasifikasikan yang benar. Selain itu, nilai kappa dilakukan untuk menilai model dengan mempertimbangkan adanya kesepakatan yang terjadi kebetulan. Perbandingan lima model menggunakan nilai akurasi dan kappa dapat dilihat pada tabel 10 di bawah ini.

**Tabel 10. Perbandingan Model**

No	Pembuatan Material dari Penelitian		
	Model	Overall Accuracy(%)	Value Kappa(%)
1	<i>Random Forest</i>	91,34%	82,40%
2	<i>Decision Tree</i>	89,76%	79,58%
3	<i>K-Nearest Neighbor</i>	91,34%	82,67%
4	<i>Support Vector Machine</i>	93,70%	87,36%
5	<i>Naïve Bayes</i>	87,40%	74,82%

Berdasarkan tabel 10 di atas dapat diketahui bahwa model *Random Forest* mengatasi variabilitas data menggunakan gabungan prediksi dari banyak pohon keputusan. Model ini memiliki nilai akurasi sebesar 91%, maka model sangat baik dalam melakukan pengklasifikasian yang benar. Selain itu, terdapat nilai kappa sebesar 82%, menunjukkan model dapat mengatasi ketidakseimbangan kelas dan memberikan prediksi lebih seimbang di antara kelas-kelas yang berbeda dengan sangat baik.

Kemudian model *Decision Tree* mengatasi variabilitas data dengan membagi data menjadi sub bagian yang lebih kecil. Model ini memiliki nilai akurasi sebesar 90% yang berarti bahwa model baik dalam melakukan pengklasifikasian yang benar. Selain itu, terdapat nilai kappa sebesar 80%, berarti



bahwa model yang didapatkan mampu mengatasi ketidakseimbangan kelas dan memberikan prediksi lebih seimbang di antara kelas-kelas yang berbeda dengan baik.

Selain itu, model *K-Nearest Neighbor* mengatasi variabilitas data dengan memilih jumlah tetangga terdekat yang optimal. Model ini memiliki nilai akurasi sebesar 91%. Hal tersebut menunjukkan bahwa model sangat baik dalam melakukan pengklasifikasian yang benar. Kemudian, nilai kappa yang didapatkan yaitu sebesar 83%, menunjukkan model dapat mengatasi ketidakseimbangan kelas dan memberikan prediksi lebih seimbang di antara kelas-kelas yang berbeda dengan sangat baik.

Sementara itu, model *Support Vector Machine* menggunakan fungsi kernel, yaitu memetakan data ke ruang dimensi tinggi. Model ini memiliki nilai akurasi sebesar 94% yang berarti bahwa model sangat baik dalam melakukan pengklasifikasian yang benar. Selain itu, nilai kappa yang didapatkan yaitu sebesar 87% yang berarti bahwa model dapat mengatasi ketidakseimbangan kelas dan memberikan prediksi lebih seimbang di antara kelas-kelas yang berbeda dengan sangat baik.

Selanjutnya, model *Naïve Bayes* menggunakan asumsi independensi untuk mengatasi variabilitas data. Model ini memiliki nilai akurasi sebesar 87%, maka model tersebut baik dalam melakukan pengklasifikasian yang benar. Selain itu, terdapat nilai kappa sebesar 75%, menunjukkan model dapat mengatasi ketidakseimbangan kelas dan memberikan prediksi lebih seimbang di antara kelas-kelas yang berbeda dengan baik. Hal ini juga menunjukkan model *Naïve Bayes* dapat mengklasifikasikan data yang benar dan dapat mengatasi ketidakseimbangan kelas dengan baik.

Metode terbaik dapat diketahui melalui nilai akurasi dan nilai kappa tertinggi. Dari kelima metode yang dilakukan, metode klasifikasi *Support Vector Machine* (SVM) merupakan metode klasifikasi terbaik dengan nilai akurasi dan kappa paling tinggi serta berkategori sangat baik. Maka, metode *Support Vector Machine* (SVM) paling baik digunakan untuk memprediksi *land surface temperature* di Jawa Barat.

### III.2. Pembahasan

Berdasarkan hasil penelitian mengenai suhu permukaan tanah yang dipengaruhi oleh NDBI, NDVI, NDWI, maka didapatkan model terbaik yaitu *Support Vector Machine* (SVM) dengan nilai akurasi dan nilai kappa tertinggi. Hal ini sejalan dengan penelitian yang dilakukan oleh Noi & Kappas [20] yang melakukan penelitian mengenai lingkungan yaitu penggunaan atauutupan lahan yang menyatakan hasil bahwa model dengan akurasi tertinggi yaitu algoritma SVM. Selain itu, SVM ini dianggap sebagai metode paling terkenal karena kinerjanya yang kuat pada berbagai masalah seperti klasifikasi, estimasi, regresi, pengenalan pola, dan peramalan [21]. Kemudian menurut penelitian yang dilakukan oleh Bakar et al [22], metode SVM paling efisien dalam melakukan klasifikasi dibandingkan Maximum Likelihood (MLC). Pada penelitian tersebut digunakan variabel NDVI, NDBI dan MNDWI untuk melihat hubungan antara LST dan *Land Use/Land Cover*.

Hal tersebut tidak terlepas dari keunggulan metode SVM dibandingkan metode yang lain. Menurut Cristianini et al [23], metode SVM menggunakan fungsi kernel yang digunakan untuk memetakan data non-linear ke ruang dimensi yang lebih tinggi sehingga data akan mudah dipisahkan secara linear. Kemudian *hyperplane* yang digunakan untuk pemisah mampu memberikan margin yang maksimal sehingga model SVM menjadi lebih robust terhadap outlier. Metode SVM sangat cocok digunakan untuk melakukan klasifikasi citra satelit terutama di bidang lingkungan karena memiliki kemampuan generalisasi yang baik dan efektif [24].



#### IV. KESIMPULAN

Provinsi Jawa Barat memiliki proporsi kecamatan dengan suhu permukaan tanah yang cukup seimbang antara yang berada dalam kategori normal dan tidak normal. Kecamatan yang memiliki suhu permukaan tanah yang tidak normal cenderung mengelompok dan merupakan wilayah metropolitan. Kemudian dengan menggunakan dataset LST, NDBI, NDVI, NDWI yang digunakan untuk membandingkan lima metode klasifikasi yaitu *Random Forest*, *Decision Tree*, *K-Nearest Neighbour*, *Support Vector Machine*, *Naïve Bayes* didapatkan model klasifikasi yang terbaik diberikan oleh metode SVM. Hal tersebut dikarenakan metode SVM memberikan hasil yang sangat baik dengan nilai akurasi tertinggi sebesar 93,70 persen dan nilai kappa tertinggi sebesar 87,36 persen. Oleh karena itu, metode *Support Vector Machine* (SVM) dapat digunakan sebagai acuan untuk melakukan prediksi.

#### REFERENSI

1. WMO. (2023). Provisional State of the Global Climate 2023.
2. IPCC. (2023). Sections. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, hal, 35-115, doi: 10.59327/IPCC/AR6-9789291691647 .
3. BMKG. (2023). Climate Outlook: Pandangan Iklim 2024. Jakarta.
4. BPS Provinsi Jawa Barat. (2024). Jawa Barat Dalam Angka 2024. Vol 49. Bandung.
5. Wei Zhao, Si-Bo Duan, Ainong Li, Gaofei Yin. (2019). A practical method for reducing terrain effect on land surface temperature using random forest regression, *Remote Sensing of Environment*, Volume 221, hal 635-649, ISSN 0034-4257, <https://doi.org/10.1016/j.rse.2018.12.008>
6. W. Li, L. Ni, Z. -L. Li, S. -B. Duan and H. Wu, "Evaluation of Machine Learning Algorithms in Spatial Downscaling of MODIS Land Surface Temperature," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2299-2307, July 2019, doi: 10.1109/JSTARS.2019.2896923
7. Shafizadeh-Moghadam, H., Weng, Q., Liu, H., & Valavi, R. (2020). Modeling the spatial variation of urban land surface temperature in relation to environmental and anthropogenic factors: a case study of Tehran, Iran. *GIScience & Remote Sensing*, 57(4), 483–496. <https://doi.org/10.1080/15481603.2020.1736857>
8. S. Y. Joko Prasetyo, K. Dwi Hartomo, M. C. Paseleng, D. Widiyanto Candra and B. H. Simanjuntak. (2019). "The Machine Learning to Detect Drought Risk in Central Java Using Landsat 8 OLI Remote Sensing Images," *2019 5th International Conference on Science and Technology (ICST)*. Yogyakarta. Indonesia. 1-6, doi: 10.1109/ICST47872.2019.9166197.
9. Vermote, E. F., & Kotchenova, S. (2008). "Atmospheric correction for the monitoring of land surfaces." *Journal of Geophysical Research: Atmospheres*, 113(D23).
10. Talarosha, B. (2005). MENCiptakan KENYAMANAN THERMAL DALAM BANGUNAN. In *Jurnal Sistem Teknik Industri* (Vol. 6, Issue 3).
11. Rouse, J. W., Hass, R. H., Schell, J. A., Deering, D. W., & Harlan, J. C. (1974). Monitoring the Vernal Advancement and Retrogradation (Greenwave Effect) of Natural Vegetation. NASA/GSFC Type III Final Report. Greenbelt, MD: NASA/ GSFC.
12. Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583–594. <https://doi.org/10.1080/01431160304987>
13. McFeeters, S.K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *\*International Journal of Remote Sensing\**, 17(7), 1425-1432.
14. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32
15. Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
16. Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185.
17. Suthaharan, S. (2016) Support Vector Machine. In: *Machine Learning Models and Algorithms for Big Data Classification*. Springer, Boston, 207-235. [https://doi.org/10.1007/978-1-4899-7641-3\\_9](https://doi.org/10.1007/978-1-4899-7641-3_9)
18. Yao, Y., & Zhou, B. (2010). *Naive Bayesian Rough Sets*. RSKT 2010, LNAI 6401, pp. 719–726, 2010.



19. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318. <https://doi.org/10.1016/j.compag.2018.01.009>
20. Thanh Noi, P., & Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel, Switzerland)*, 18(1). <https://doi.org/10.3390/s18010018>
21. Yaghoubzadeh-Bavandpour, A., Rajabi, M., Nozari, H., Ahmad, S. (2022). Support Vector Machine Applications in Water and Environmental Sciences. In: Bozorg-Haddad, O., Zolghadr-Asli, B. (eds) Computational Intelligence for Water and Environmental Sciences. Studies in Computational Intelligence, vol 1043. Springer, Singapore. [https://doi.org/10.1007/978-981-19-2519-1\\_14](https://doi.org/10.1007/978-981-19-2519-1_14)
22. Bakar, Suzana Binti Abu, et al. "Spatial assessment of land surface temperature and land use/land cover in Langkawi Island." *IOP conference series: Earth and environmental science*. Vol. 37. No. 1. IOP Publishing, 2016.
23. Cristianini, N., & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press.
24. Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247-259.
25. Xu, S., Cheng, J., & Zhang, Q. (2021). A random forest-based data fusion method for obtaining all-weather land surface temperature with high spatial resolution. *Remote Sensing*, 13(11), 2211.
26. Kim, M., Kim, D., & Kim, G. (2022). Examining the relationship between land use/land cover (LULC) and land surface temperature (LST) using explainable artificial intelligence (XAI) models: a case study of Seoul, South Korea. *International Journal of Environmental Research and Public Health*, 19(23), 15926.
27. Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18(1), 18.
28. Shafi, M., Jain, A., & Zaman, M. (2022, November). Applying machine learning algorithms on urban heat island (uhi) dataset. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 3* (pp. 725-732). Singapore: Springer Nature Singapore.