

# Pemetaan Lahan Impervious Bogor Menggunakan Metode PCA-SVM

Stephen Adhikurnia <sup>1)</sup>, Dyah Erny Herwindiati <sup>2)</sup>, Janson Hendryli <sup>3)</sup>

<sup>1)2)3)</sup> Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara  
Jl. Letjen S. Parman No.1, Jakarta 11440 Indonesia  
email: [stephen.535200056@stu.untar.ac.id](mailto:stephen.535200056@stu.untar.ac.id) <sup>1)</sup>, [dyahh@fti.untar.ac.id](mailto:dyahh@fti.untar.ac.id) <sup>2)</sup>, [jansonh@fti.untar.ac.id](mailto:jansonh@fti.untar.ac.id) <sup>3)</sup>

## ABSTRACT

Bogor adalah daerah penyangga air utama bagi Jakarta. Lahan terbuka hijau dan area resapan air yang melimpah di Bogor memiliki peranan penting sebagai penampung dan sumber air, terutama untuk wilayah sekitarnya, termasuk Jakarta. Namun, di tengah posisi Bogor sebagai daerah resapan, sumber serta cadangan air bagi Jakarta, luas lahan hijau di wilayah Bogor semakin berkurang setiap tahunnya. Hal ini tidak lepas dari berbagai faktor non alami yang terjadi seperti peningkatan jumlah penduduk secara signifikan dan tingkat pembangunan pemukiman serta daerah wisata. Untuk itu, diperlukan sebuah cara untuk melakukan pemetaan terhadap daerah impervious di Bogor. Lahan akan diklasifikasikan untuk tipe pervious, hijau sebagian, dan secara khusus impervious. Metode yang digunakan dalam penelitian ini adalah dan Principal Component Analysis – Support Vector Machine (SVM). Hasil akhir klasifikasi menunjukkan angka Pervious F1-Score sebesar 0.82, Hijau Sebagian F1-Score sebesar 0.73, dan Impervious F1-Score sebesar 0.87.

## Kata kunci

*Principal Component Analysis, Support Vector Machine, Kernel Gaussian, Remote Sensing*

## 1. Pendahuluan

Bogor, daerah yang sering dijuluki sebagai "Kota Hujan", terletak di Indonesia dan memiliki karakteristik geografis yang sangat krusial bagi ibukota Jakarta. Dengan curah hujan yang tinggi (2.000 hingga 3.000 mm per tahun) [1], Bogor adalah daerah penyangga air utama bagi Jakarta. Lahan terbuka hijau dan area resapan air yang melimpah di Bogor memiliki peranan penting sebagai penampung dan sumber air, terutama untuk wilayah sekitarnya, termasuk Jakarta. Wilayah ini memiliki banyak lahan hijau yang luas, yang berfungsi sebagai lapisan penyerap air. Aliran sungai utama seperti Sungai Ciliwung dan Sungai Cisadane memiliki peran penting dalam menyediakan pasokan air untuk berbagai kebutuhan perkotaan, pertanian, dan industri di Jakarta.

Selain itu, Bogor juga merupakan sumber mata air alami yang menjadi pemasok air terbesar serta cadangan

air bersih bagi wilayah Jakarta dan sekitarnya. Menurut catatan sejarah, pencarian sumber mata air utama bagi Jakarta dimulai pada tahun 1918 oleh Pemerintah Hindia-Belanda. Pencarian awal dimulai di daerah sekitar kaki Gunung Salak hingga ditemukanlah sumber mata air strategis di Ciomas yang dapat menghasilkan debit air sebanyak 484 liter air bersih per detik. Pembangunan sarana dan prasarana di wilayah setempat dimulai sejak tahun 1920. Hasil penemuan sumber air bersih tersebut dialirkan melalui prasarana yang telah dibangun hingga sampai ke wilayah Jakarta. Pada tahun 1922, dibentuklah sebuah badan usaha milik daerah DKI Jakarta yaitu Perusahaan Umum Daerah Air Minum Jaya (PAM JAYA) untuk menaungi dan mengawasi pasokan air bersih dari Bogor ke Jakarta hingga saat ini [2].

Namun, di tengah posisi Bogor sebagai daerah resapan, sumber serta cadangan air bagi Jakarta, luas lahan hijau di wilayah Bogor semakin berkurang setiap tahunnya. Hal ini tidak lepas dari berbagai faktor non alami yang terjadi. Menurut data dari Badan Pusat Statistik, rata-rata jumlah penduduk di Bogor meningkat sebesar 1,37% per tahun [3]. Jumlah penduduk yang semakin meningkat baik secara alami maupun melalui urbanisasi kemudian menyebabkan tingkat kepadatan penduduk semakin tinggi yang pada akhirnya akan menyebabkan meningkatnya angka alih fungsi lahan hijau.

Selain itu, tingkat pembangunan pemukiman dan daerah wisata di Bogor juga meningkat setiap tahunnya yang memperbesar luas daerah lahan impervious. Lahan impervious atau lahan tidak hijau merupakan tipe lahan yang tidak dapat menyerap air atau menampung air hujan di dalam tanahnya seperti jalanan, trotoar, bangunan, serta material lainnya yang tidak dapat menyerap atau menampung air. Menurut Badan Penanggulangan Bencana Daerah (BPBD) Bogor, dalam rentang 1 tahun terakhir, Bogor tak lepas dari bencana kekeringan yang melanda setidaknya 8 Kecamatan [4]. Hal ini menunjukkan bahwa daerah cadangan air di Bogor semakin berkurang dan lahan impervious semakin luas seiring berjalannya waktu.

Maka daripada itu, tulisan ilmiah ini bertujuan untuk memberikan informasi pemetaan terhadap lahan hijau, hijau sebagian, dan secara khusus lahan impervious

untuk setiap daerah kecamatan di Bogor. Salah satu metode untuk dapat diandalkan untuk melakukan analisis pemetaan terhadap kondisi lahan di Bogor adalah dengan menggunakan bantuan citra satelit. Pada skripsi ini, dengan bantuan data citra satelit, akan dilakukan proses pemetaan terhadap lahan hijau, hijau sebagian, dan *impervious*. Data masukan yang digunakan merupakan data citra satelit *landsat 8* untuk band 2 (Blue), 3(Green), 4(Red), 5(Near-Infrared), 6(SWIR 1), dan 7(SWIR 2) yang diambil dan dipotong secara spesifik untuk wilayah Bogor.

Metode yang digunakan dalam penelitian ilmiah ini adalah Support Vector Machine dan Principal Component Analysis - Support Vector Machine (PCA-SVM). Kedua jenis metode ini akan dilakukan analisis untuk membandingkan tingkat akurasi antar kedua metode tersebut.

Support Vector Machine (SVM) merupakan metode untuk klasifikasi yang didasarkan pada prinsip maksimisasi margin. SVM melakukan minimisasi risiko struktural, yang meningkatkan kompleksitas klasifikasi dengan tujuan mencapai performa generalisasi yang baik. Pada penelitian ini akan digunakan kernel jenis Gaussian dikarenakan kernel Gaussian memiliki kecenderungan untuk lebih *robust* atau tahan terhadap *noise* dalam data serta memiliki keunggulan dalam mengolah data dengan jumlah fitur yang besar.

PCA-SVM merupakan kombinasi dari dua metode, yaitu Principal Component Analysis (PCA) dan Support Vector Machine (SVM). PCA merupakan suatu metode statistik yang digunakan untuk mengurangi dimensi variabel pada dataset dengan mempertahankan sebanyak mungkin informasi yang relevan. PCA memiliki manfaat yang signifikan dalam mengatasi masalah multikolinearitas pada dataset yang digunakan. PCA digunakan dalam proses pengolahan data untuk penelitian ini dikarenakan data *remote sensing* memiliki kecenderungan untuk memiliki sifat multikolinear. Multikolinearitas adalah sebuah kondisi yang terjadi ketika terdapat hubungan linear yang kuat antara dua atau lebih variabel independen dalam suatu model. Jika hubungan linear antar variabel independen sebagaimana disebutkan sebelumnya merupakan korelasi yang sempurna, maka variabel tersebut dapat dikatakan berkolinearitas ganda sempurna. Korelasi yang terlampaui kuat akan meningkatkan redundansi informasi. Dengan mengurangi dimensi variabel pada dataset dan menghasilkan komponen utama yang tidak berkorelasi satu sama lain, PCA dapat membantu mengatasi masalah multikolinearitas pada data *remote sensing* yang digunakan. Selain itu, proses reduksi dimensi pada proses PCA akan mengurangi waktu komputasi pada saat dilakukan proses klasifikasi berikutnya dengan Support Vector Machine.

Penelitian sebelumnya terkait pemetaan lahan daerah Bogor dan penggunaan metode Support Vector Machine yang berjudul “Performa Support Vector Machine Pada Klasifikasi Lahan dan Air Tanah” menghasilkan klasifikasi terhadap 3 jenis lahan yaitu *pervious*, hijau

sebagian, dan *impervious* serta membuktikan bahwa model Support Vector Machine dengan kernel linear memiliki performa yang lebih baik dibandingkan dengan kernel polinomial [5].

Adapun penelitian terkait pemetaan lahan yang berjudul “Kajian Ketersediaan Air Tanah Terkait Pemanfaatan Lahan di Kabupaten Blitar” yang memanfaatkan analisis infiltrasi untuk menentukan cadangan air pada tanah [6]. Selain itu, penelitian yang berjudul “Land Use Change Using Least Absolute Shrinkage and Selection Operator Regression in Jakarta's Buffer Cities” yang memanfaatkan metode Least Absolute Shrinkage and Selection Operator (LASSO) dalam melakukan pemetaan lahan [7].

## 2. Landasan Teori

Untuk mendukung kajian ilmiah yang dibahas, ada beberapa dasar teori yang perlu untuk diperhatikan dan dipahami.

### 2.1. Lokasi dan Waktu Penelitian

Penelitian ini mengambil sampel wilayah Bogor sebagai batasan area kajian. Bogor merupakan sebuah daerah terletak di provinsi Jawa Barat, Indonesia. Daerah Bogor membentang dari 6°18' hingga 6°47'10 Lintang Selatan dan 106°23'45 hingga 107° 13'30 Bujur Timur. Bogor berbatasan dengan Kabupaten Tangerang (Banten), Kota Depok, Kota Bekasi, dan Kabupaten Bekasi di utara, Kabupaten Karawang di timur, Kabupaten Cianjur dan Kabupaten Sukabumi di selatan, serta Kabupaten Lebak (Banten) di barat.

Daerah Bogor pada penelitian ini dapat dibagi menjadi Kota Bogor dan Kabupaten Bogor. Pada penelitian ini, ruang lingkup yang menjadi batasan penelitian adalah seluruh 40 kecamatan yang ada di Kabupaten dengan ditambah Kota Bogor yang tidak dipecah menjadi perkecamatan yang dapat dilihat pada Gambar berikut :



Gambar 1 Peta Kabupaten dan Kota Bogor

### 2.2. Satelit Landsat 8

Satelit Landsat 8 adalah sebuah satelit observasi bumi milik The National Aeronautics and Space Administration (NASA) dan United States Geological

Survey (USGS) yang diluncurkan pada tanggal 11 Februari 2013. Satelit Landsat 8 merupakan salah satu versi terbaru dalam satelit seri Landsat serta dapat menghasilkan pandangan yang lebih jelas dan lebih baik untuk pengamatan resolusi spasial dibandingkan kebanyakan instrumen penginderaan lainnya dan sensitivitas yang lebih besar terhadap warna dan tingkat kecerahan dibandingkan satelit Landsat versi sebelumnya.

Satelit Landsat 8 memiliki dua buah sensor yaitu Operational Land Imager (OLI) dan Thermal Infrared Sensor (TIRS). Sensor OLI yang terdiri dari band 1, 2, 3, 4, 5, 6, 7, 8, dan 9 menghasilkan data citra multispektral dengan resolusi spasial 30 meter. Sensor TIRS yang terdiri dari band 10 dan 11 berfungsi untuk memberikan data inframerah terkait panas dengan resolusi spasial 100 meter.

Berikut merupakan data seluruh band yang terdapat pada Satelit Landsat 8 beserta jangkauan panjang gelombang dan kegunaannya :

Tabel 1 Band Pada Satelit Landsat 8

Band	Panjang Gelombang (Mikrometer)	Penggunaan dalam Pemetaan
Band 1 – Coastal Aerosol	0.43-0.45	Air
Band 2 – Blue	0.45-0.51	Pemetaan batimetri, membedakan tanah dari tumbuh-tumbuhan
Band 3 – Green	0.53-0.59	Mempertegas puncak vegetasi
Band 4 – Red	0.64-0.67	Membedakan lereng/slope vegetasi
Band 5 – Near Infrared (NIR)	0.85-0.88	Mempertegas kandungan biomassa dan garis pantai
Band 6 – Short-wave Infrared (SWIR) 1	1.57-1.65	Membedakan kadar air tanah dan tumbuh-tumbuhan serta dapat menembus awan yang tipis
Band 7 – Short-wave Infrared (SWIR) 2	2.11-2.29	Meningkatkan penggambaran kadar air tanah dan tumbuh-tumbuhan serta dapat menembus awan yang tipis
Band 8 - Panchromatic	0.50-0.68	Menghasilkan gambar yang lebih tajam
Band 9 – Cirrus	1.36-1.38	Meningkatkan kemampuan deteksi kontaminasi awan cirrus
Band 10 – TIRS 1	10.60-11.19	Pemetaan panas dan kelembapan tanah
Band 11 –	11.5-12.51	Pemetaan panas dan

TIRS 2		kelembapan tanah
--------	--	------------------

### 2.3. Principal Component Analysis

Principal Component Analysis (PCA) merupakan sebuah metode untuk mengidentifikasi pola dalam data, serta mengekspresikan data sedemikian rupa untuk menyoroti apa yang menjadi persamaan dan perbedaan dalam data [8]. Karena pola dalam data akan sulit untuk ditemukan dalam data dimensi tinggi yang dimana representasi grafis tidak tersedia, PCA menjadi sebuah alat alternatif yang ampuh untuk melakukan analisis data. PCA dapat dimanfaatkan untuk menganalisis data dalam skala besar. PCA mengolah data dalam skala besar dengan cara mereduksi dimensi data namun tetap mempertahankan informasi-informasi yang penting. PCA dapat membantu mengatasi data yang memiliki sifat multikolinearitas yang tinggi. Multikolinearitas sendiri merupakan sebuah kondisi yang terjadi ketika terdapat hubungan linear yang kuat antara dua atau lebih variabel independen dalam suatu model. Multikolinearitas dapat berdampak pada koefisien algoritma klasifikasi yang menjadi tidak stabil sehingga diperlukan proses reduksi dimensi untuk mengatasi masalah multikolinearitas.

Hasil akhir yang diharapkan dari penggunaan metode PCA adalah berupa transformasi dari dataset asli ke ruang fitur baru yang disebut sebagai "komponen utama". Data akan ditransformasikan kembali ke bentuk semula namun dengan tingkatan dimensi yang lebih rendah. PCA dimulai dengan proses normalisasi data. Tujuan utama yang ingin dicapai adalah agar data memiliki mean nol dan variansi yang seragam untuk setiap fiturnya. Selain itu, metode PCA juga cukup sensitif terhadap nilai variansi variabel awal sehingga data perlu untuk dinormalisasi terlebih dahulu. Hal ini juga akan membantu untuk menghilangkan bias pada hasil akhir yang disebabkan oleh dominasi rentang variabel yang lebih luas dibandingkan dengan rentang variabel lainnya. Adapun rumus untuk melakukan normalisasi data adalah sebagai berikut :

$$X_{std} = \frac{X - \mu}{\sigma} \tag{1}$$

Keterangan:

Xstd = Hasil Data yang telah distandarisasi

X = Data Asli

μ = Nilai Mean

σ = Standard Deviasi

Kemudian, hasil dari matriks yang telah distandarisasi akan dihitung dan dibentuk ke dalam matriks kovarians. Tujuan utama yang ingin dicapai adalah untuk melihat dan mengidentifikasi apakah setiap fitur dalam data memiliki hubungan antar satu dengan yang lainnya. Matriks kovarians sendiri merupakan sebuah matriks simetris dengan dimensi yang sama. Rumus perhitungan kovarians antar 2 variabel adalah sebagai berikut :

$$Cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{x})(Y_i - \bar{y}) \tag{2}$$

Keterangan :

- Cov(X, Y) = kovarians antara variabel acak X dan Y.
- n = jumlah observasi atau data yang kita miliki.
- I = indeks yang digunakan dalam penjumlahan.
- X<sub>i</sub> = nilai dari variabel acak X pada observasi ke-i.
- $\bar{X}$  (X bar) = rata-rata (mean) dari semua nilai X.
- Y<sub>i</sub> = nilai dari variabel acak Y pada observasi ke-i.
- $\bar{Y}$  (Y bar): rata-rata (mean) dari semua nilai Y.

Eigen vektor dan eigen value adalah prinsip dari aljabar linear yang digunakan untuk mengkalkulasi hasil dari matriks kovarians. Eigen value merupakan skalar untuk menjelaskan variasi dalam data. Adapun rumus perhitungan eigen value adalah sebagai berikut :

$$\det(A - \lambda I) = 0 \tag{3}$$

Keterangan :

- det = Determinan
- A = Matriks Kovarians
- I = Matriks Identitas
- $\lambda$  = Nilai Eigen Value

Sedangkan Eigen vektor merupakan jenis vektor yang tidak akan berubah arahnya meskipun terdampak oleh transformasi linear. Adapun rumus perhitungan Eigen vektor adalah sebagai berikut :

$$(A - \lambda I)v = 0 \tag{4}$$

Keterangan :

- A = Matriks Kovarians
- I = Matriks Identitas
- $\lambda$  = Nilai Eigen Value
- v = Nilai Eigen Vektor

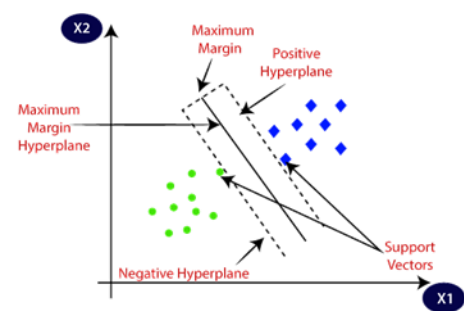
Nilai Eigen Value akan diurutkan dari yang terbesar hingga terkecil (descending order). Eigen Vektor yang memiliki nilai Eigen Value terbesar akan menjadi komponen pertama. Begitupun seterusnya untuk pengurutan komponen disesuaikan pada besar Eigen Valuenya. Pemilihan jumlah komponen utama yang diambil bersifat fleksibel sesuai dengan kebutuhan penelitian ataupun yang dianggap paling baik. Semakin tinggi nilai Eigen Value pada tahapan sebelumnya, maka kontribusi komponen utama tersebut terhadap variasi data juga semakin besar. Hal dasar yang perlu dipertimbangkan dalam memilih jumlah komponen adalah tetap mempertahankan variasi data dengan di sisi lain juga tetap dapat mereduksi dimensi data.

### 2.4. Support Vector Machine

Support Vector Machine (SVM) merupakan sebuah algoritma klasifikasi supervised yang dapat digunakan untuk mempelajari pola data dengan tingkat akurasi dan reproduktifitas yang tinggi. SVM merupakan metode

untuk klasifikasi yang didasarkan pada prinsip maksimisasi margin [9]. Pada dasarnya, konsep utama dari algoritma SVM adalah dengan mencari hyperplane atau fungsi pemisah yang paling baik untuk memisahkan beberapa kelas atau label. SVM memiliki keunggulan untuk bekerja dengan ruang dimensi yang tinggi. Fungsi matematis utama pada algoritma SVM adalah kernel. Kernel merupakan metode perhitungan yang digunakan dalam proses pemetaan ruang dimensi. Hal ini memungkinkan bagi metode SVM untuk dapat mengatasi problema pemisahan data yang tidak linier dalam bentuk fungsi pemisah di ruang fitur yang lebih tinggi.

Selain keunggulan dalam melakukan pemisahan data pada ruang dimensi yang tinggi, SVM juga memiliki tingkat fleksibilitas yang sangat toleran untuk menangani berbagai jenis masalah klasifikasi, termasuk masalah klasifikasi biner dan multi-kelas yang dimana di dalam penelitian ini akan menggunakan pendekatan multi-kels. Dalam masalah klasifikasi biner, SVM berusaha untuk memisahkan dua kelas, sedangkan dalam masalah klasifikasi multi-kelas, SVM dapat diterapkan dengan beberapa pendekatan, termasuk satu lawan semua (one-vs-all) di mana setiap kelas dibedakan dari yang lainnya. Dalam pendekatan one-vs-all, setiap kelas akan didefinisikan sebagai satu kelas positif, sementara semua kelas lainnya dianggap sebagai kelas negatif. Setiap SVM memisahkan satu kelas positif dari kelas negatif (kelas yang lainnya). Pengklasifikasian data baru akan memproses dan menguji data tersebut pada semua data dan kelas dengan hasil yang paling tinggi (nilai positif maksimum) dipilih sebagai hasil klasifikasi. Adapun konsep dasar dari metode SVM dapat dilihat pada gambar berikut :



Gambar 2 Konsep Support Vector Machine (SVM)

Pada gambar konsep SVM diatas, dapat diambil kesimpulan bahwa SVM memiliki maximum margin yang merupakan upaya untuk mencari pemisah (hyperplane) antara dua kelas data yang memiliki margin terbesar. Margin adalah jarak antara hyperplane dan titik terdekat dari masing-masing kelas data. Selain itu terdapat juga maximum margin hyperplane yang merupakan hyperplane yang dipilih dalam model SVM yang memiliki margin terbesar. Hyperplane pada maximum margin hyperplane adalah pemisah optimal antara dua kelas data. Batas atas dan batas bawah dari hyperplane akan didefinisikan sebagai positive

hyperplane dan negative hyperplane. Positive hyperplane merupakan jenis hyperplane yang mendefinisikan batas atas dari kelas positif sedangkan negative hyperplane merupakan jenis hyperplane yang mendefinisikan batas atas dari kelas negatif. Titik-titik data yang berada paling dekat dengan hyperplane pemisah atau maximum margin hyperplane dapat dikategorikan sebagai support vectors. Support vectors adalah titik-titik yang memiliki peran signifikan dalam menentukan posisi dan orientasi hyperplane.

### 2.5. Kernel Trick

Kernel trick adalah sebuah teknik penting dalam pembelajaran mesin, terutama dalam konteks Support Vector Machine (SVM). Kernel trick memungkinkan pemodelan non-linear data dengan lebih efisien. Saat bekerja dengan data yang tidak dapat dipisahkan secara linear di ruang fitur asli, kernel trick akan membantu untuk mentransformasikan data ke dalam ruang fitur yang lebih tinggi dimensi sehingga pemisahan linear menjadi mungkin. Hal dilakukan tanpa perlu menghitung semua titik data baru di ruang fitur yang lebih tinggi, sehingga menghemat waktu dan sumber daya komputasi.

Dalam penelitian ini, kernel trick yang akan digunakan adalah Gaussian atau Radial Basis Function. Kernel Gaussian bekerja dengan cara mengukur sejauh mana setiap titik data dari titik referensi atau pusat kluster yang dihitung sebagai titik referensi. Adapun persamaan fungsi kernel gaussian adalah sebagai berikut :

$$K(x_i, x_j) = \exp(-\gamma ||x_i - x_j ||^2) \tag{5}$$

Keterangan :

$x_i$  &  $x_j$  = dua titik/vektor data latih

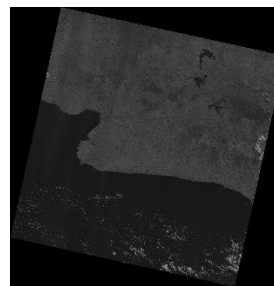
$\exp$  = eksponensial

$\gamma$  = gamma (parameter kernel)

## 3. Hasil Percobaan

### 3.1. Menyiapkan Dataset

Sumber dataset yang digunakan dalam penelitian ini diperoleh dari situs <https://earthexplorer.usgs.gov/> milik Departemen Dalam Negeri Amerika Serikat untuk satelit Landsat 8 dan hanya diambil untuk band 2 (Blue), 3(Green), 4(Red), 5(Near-Infrared), 6(SWIR 1), dan 7(SWIR 2). Dari segi lingkup penggunaan data, wilayah yang dianalisis hanya untuk daerah Bogor yang terdiri dari 40 kecamatan di Kabupaten Bogor dan 1 Kota Bogor. Pemilihan daerah Bogor sebagai wilayah untuk dilakukan penelitian didasari pada sumber dan cadangan air bersih utama bagi Jakarta berasal dari Bogor. Aliran air utama dari kaki Gunung Salak akan mengalir mengikuti arus dengan sumber aliran utama yaitu Sungai Ciliwung dan sumber-sumber pendamping lainnya. Contoh data citra yang diunduh dari situs <https://earthexplorer.usgs.gov/> adalah sebagai berikut :



Gambar 3 Citra Satelit Band 5 Daerah Bogor Tahun 2020

#### 3.1.1. Pra-Pemrosesan Data

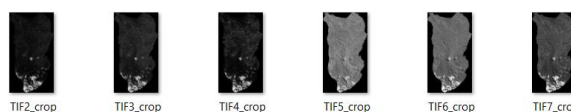
Tahap pra-pemrosesan data pertama kali dilakukan menggunakan bantuan perangkat lunak ArcMap. ArcMap merupakan sebuah perangkat lunak yang membantu mentransformasikan data berukuran 16 bit menjadi 8 bit. Penggunaan bit yang lebih rendah dapat meningkatkan kecepatan dalam proses pengolahan data.

Kemudian, data akan dilakukan proses pemotongan untuk tiap-tiap batasan kecamatan dengan menggunakan perangkat lunak Environment for Visualizing Images (ENVI). Pada perangkat lunak ENVI, dilakukan penginputan data-data batas wilayah untuk tiap kecamatan yang didapatkan dengan bantuan perangkat Google Earth Pro. Data setelah dilakukan pemotongan akan terdiri dari 40 set data untuk masing-masing wilayah kecamatan di Kabupaten Bogor dan 1 set data Kota Bogor. Data tidak dilakukan pra-proses untuk menghapus awan yang ada pada tangkapan citra. Dari data hasil pra-pemrosesan, digunakanlah 1 sampel kecamatan untuk tiap-tiap jenis lahan baik pervious, hijau Sebagian, dan impervious yang dapat dilihat pada tabel berikut :

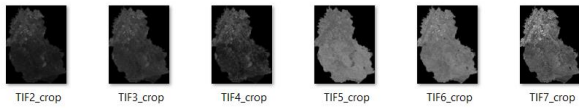
Tabel 2 Sampel Dataset yang Digunakan

Data Latih	Jenis Lahan		
	Hijau	Hijau Sebagian	Impervious
Wilayah	Sukajaya	Caringin	Kota Bogor

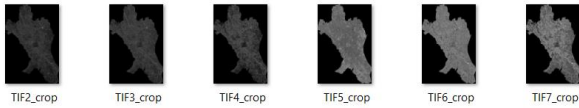
Pengambilan sampel data ini didasari oleh kondisi sebenarnya pada wilayah tercantum. Kecamatan Sukajaya didominasi oleh lahan hijau yang terdiri dari area pepohonan taman nasional dan perkebunan. Kecamatan Citeureup memiliki area pemukiman dan lahan hijau yang relatif berimbang. Sedangkan untuk Kota Bogor memiliki tingkat pemukiman yang sangat luas dan padat. Berikut merupakan gambar citra yang telah dilakukan pra-pemrosesan untuk masing-masing sampel wilayah :



Gambar 4 Dataset Citra B2 – B7 Kecamatan Sukajaya

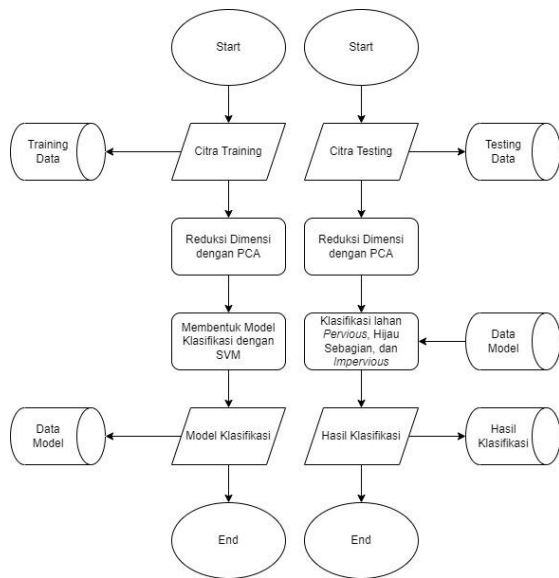


Gambar 5 Dataset Citra B2 – B7 Kecamatan Citeureup



Gambar 6 Dataset Citra B2 – B7 Kota Bogor

Setelah tahap pra-pemrosesan selesai dilakukan, data citra akan diolah untuk membangun model klasifikasi terbaik di tahap pelatihan dan untuk menguji model tersebut di tahap pengujian. Secara garis besar, alur pengujian dan pelatihan dapat dilihat pada gambar berikut :



Gambar 7 Flowchart Pelatihan dan Pengujian

### 3.1.2. Pembacaan Data

Proses pengolahan data setelah dilakukan pra-pemrosesan sebelumnya dengan menggunakan ENVI adalah dengan membaca data gambar masing-masing band ke dalam bentuk matriks yang berisi angka-angka. Proses ini dapat dilakukan dengan bantuan library python rasterio. Hasil pembacaan masing-masing band akan ditransformasikan kedalam vektor kolom dan digabung menjadi 1 buah matriks baru. Sebagai contoh, berikut merupakan data band untuk kecamatan Sukajaya yang berukuran 387 x 697 yang telah diolah menjadi sebuah matriks baru :

Tabel 3 Hasil Matriks Pembacaan untuk Kecamatan Sukajaya

Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

...	...	...	...	...	...
68	58	60	97	89	82
63	61	63	95	95	88
82	86	88	110	128	119
73	78	79	104	114	103
...	...	...	...	...	...
0	0	0	0	0	0

Selanjutnya setiap baris data yang mengandung nilai 0 pada seluruh kolomnya akan dihapus dan tidak dimasukkan sebagai set data karena tidak mengandung informasi. Pada penelitian ini juga dilakukan proses sampling dataset dengan mengambil masing-masing sebanyak 20.000 baris untuk tiap kategori lahan pervious, hijau sebagian, dan impervious sehingga total data latih yang digunakan adalah sebanyak 60.000 baris data.

### 3.2. Proses Pelatihan dan Pengujian

Berdasarkan pada gambar 7 terkait Flowchart Pelatihan, proses pelatihan dan pengujian data akan melalui tahapan reduksi dimensi, pembentukan model klasifikasi, serta proses klasifikasi lahan. Reduksi dimensi akan dilakukan dengan metode Principal Component Analysis. Selanjutnya data hasil reduksi dimensi akan dibagi sebagai data pelatihan dan pengujian atau validasi untuk membentuk sebuah model. Model yang telah terbentuk akan digunakan untuk melakukan klasifikasi pada daerah kecamatan di Bogor ke dalam 3 jenis lahan yaitu pervious, hijau sebagian, dan impervious.

#### 3.2.1. Reduksi Dimensi

Pada penelitian ini, metode Principal Component Analysis (PCA) dimanfaatkan untuk mereduksi dimensi dari matriks data citra. Proses ini akan mengkalulasikan nilai eigen value dan eigen vector serta mengurutkan nilai eigen dari yang terbesar hingga terkecil. Proses yang dilakukan pada data latih dan data uji hanya mempertahankan 4 komponen eigen value dan eigen vektor teratas yang memiliki kontribusi komponen terbesar terhadap variasi data.

#### 3.2.2. Pembentukan Model Klasifikasi

Pada penelitian ini, model yang akan dibentuk dan digunakan adalah model Principal Component Analysis – Support Vector Machine (PCA-SVM) sebagai model utama dan model Support Vector Machine konvensional tanpa PCA sebagai model pembanding. Kedua model tersebut akan dibangun menggunakan kernel Gaussian dan akan memiliki dataset input pelatihan yang sama. Data pelatihan dan pengujian atau validasi akan dibagi dengan rasio 8 : 2. Hal ini bertujuan untuk memberikan trade-off yang baik antara jumlah data yang digunakan untuk melatih model (80%) dan jumlah data yang digunakan untuk menguji kinerja model (20%). Dengan memiliki data pelatihan yang lebih banyak akan memungkinkan sistem dapat mempelajari lebih banyak juga untuk menghasilkan keputusan yang lebih baik. Hasil percobaan pelatihan

dan pengujian masing-masing model akan dievaluasi menggunakan bantuan confusion matrix. Convusion matrix merupakan alat bantu untuk menghitung akurasi model, nilai precision, recall, dan F1-Score. Precision mengukur sejauh mana prediksi True Positive dibandingkan dengan total prediksi positif (True Positive dan False Positive). Recall atau sensitivitas mengukur sejauh mana kasus True Positive dibandingkan dengan total jumlah kasus positif sebenarnya (True Positive dan False Negative). F1-Score adalah rata-rata harmonik dari precision dan recall yang berguna ketika distribusi kelas tidak seimbang atau kelas positif sangat penting untuk dikenali dengan akurasi yang baik. Adapun hasil perhitungan convusion matrix yang telah didapatkan untuk perbandingan performa model dapat dilihat pada tabel berikut :

Tabel 4 Perbandingan Performa Model

Parameter	SVM	PCA-SVM
Akurasi	0.7978	0.8076
Pervious Precision	0.83	0.84
Pervious Recall	0.79	0.80
Pervious F1-Score	0.81	0.82
Hijau Sebagian Precision	0.74	0.75
Hijau Sebagian Recall	0.68	0.70
Hijau Sebagian F1-Score	0.71	0.73
Impervious Precision	0.82	0.82
Impervious Recall	0.92	0.92
Impervious F1-Score	0.86	0.87
Macro Avg Precision	0.80	0.81
Macro Avg Recall	0.80	0.81
Macro Avg F1-Score	0.80	0.81

Berdasarkan hasil evaluasi antar kedua model pada tabel diatas, dapat diketahui bahwa model yang dilakukan proses reduksi dimensi dengan bantuan PCA terlebih dahulu memiliki tingkatan akurasi yang sedikit lebih baik dibandingkan dengan model yang tidak dilakukan proses reduksi dimensi.

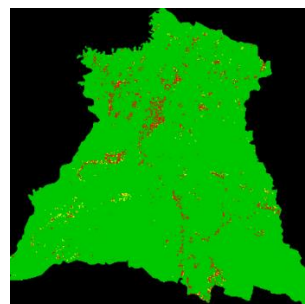
### 3.2.3. Proses Klasifikasi Lahan

Hasil dari pembentukan model pada proses sebelumnya akan digunakan untuk melakukan proses klasifikasi terhadap data uji. Model yang telah dibentuk bertujuan untuk dapat melakukan tebakan atau prediksi berdasarkan hasil pelatihan sebelumnya. Berdasarkan pada gambar 7 terkait Flowchart Pelatihan, data gambar citra yang akan diuji juga akan melalui proses reduksi dimensi untuk menyesuaikan set data pada proses pelatihan. Setiap set data gambar citra yang akan dibaca dan digunakan untuk proses klasifikasi juga akan diolah ke dalam bentuk vektor kolom yang kemudian akan digabungkan untuk setiap bandnya menjadi sebuah matriks baru. Proses penggabungan data seluruh band menjadi satu matriks baru sama seperti pada saat proses pelatihan.

Selanjutnya, dengan bantuan library scikit-image pada python, akan dibentuk sebuah gambar baru berdasarkan hasil klasifikasi pada data citra pengujian. Gambar yang akan terbentuk nantinya akan memiliki

ukuran piksel yang sama dengan data citra yang digunakan untuk proses klasifikasi. Setiap kali model menebak impervious, maka gambar baru akan diberi warna hijau, warna kuning diberikan untuk tebakan hijau Sebagian, dan warna merah digunakan untuk tebakan impervious. Untuk setiap pembacaan terhadap baris data yang memiliki nilai 0 pada band 2 sampai band 7 akan secara otomatis diberi warna hitam karena bukan termasuk cakupan wilayah yang dilakukan proses klasifikasi.

Sebagai contoh, data Kecamatan Cariu untuk bulan April 2020 akan dijadikan sampel untuk dilakukan proses klasifikasi. Hasil akhir gambar baru yang terbentuk dari proses klasifikasi untuk Kecamatan Cariu adalah sebagai berikut :



Gambar 8 Hasil Klasifikasi Kecamatan Cariu

### 3.3. Analisa dan Evaluasi Hasil

Hasil akhir dari penelitian akan ditunjukan untuk melakukan klasifikasi antara lahan impervious, hijau sebagian dan pervious untuk daerah-daerah kecamatan di Bogor. Lahan impervious akan diberi warna merah, lahan hijau sebagian akan diberi warna kuning serta lahan pervious akan diberi warna hijau. Hasil akhir klasifikasi untuk Kecamatan Cariu menunjukkan bahwa dari 106.108 ribu piksel yang diuji, didapati 100.647 piksel pervious, 477 piksel hijau Sebagian, dan 4.984 piksel impervious. Berikut merupakan gambar hasil akhir dari klasifikasi untuk daerah Kecamatan Cariu di Bogor :



Gambar 9 Perbandingan Citra Asli Google Earth dan Citra Hasil Klasifikasi Kecamatan Cariu

Pada gambar perbandingan diatas, dapat dilihat pada tampilan Google Earth bahwa citra asli Kecamatan Cariu memiliki lahan yang cenderung hijau namun juga tampak sedikit padat atau impervious pada beberapa titik. Citra hasil klasifikasi juga menampilkan bahwa

lahan Kecamatan Cariu didominasi oleh lahan pervious dengan hijau sebagian dan impervious di beberapa titik. Banyaknya lahan yang masih berwarna hijau dan sangat sedikitnya lahan berwarna merah pada Kecamatan Cariu menandakan bahwa wilayah tersebut merupakan lahan yang memiliki banyak ketersediaan pasokan dan cadangan air bersih seperti hutan, pepohonan dan lahan serapan lainnya.

#### 4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, penulis dapat menarik kesimpulan sebagai berikut :

1. Metode Principal Component Analysis – Support Vector Machine (PCA-SVM) dapat digunakan secara efektif dan akurat untuk melakukan klasifikasi pemetaan lahan dengan tingkat Pervious F1-Score sebesar 0.82, Hijau Sebagian F1-Score sebesar 0.73, dan Impervious F1-Score sebesar 0.87
2. Metode Principal Component Analysis – Support Vector Machine (PCA-SVM) memiliki akurasi model yang sedikit lebih akurat dibandingkan dengan Support Vector Machine (SVM) konvensional dengan perbandingan akurasi sebesar 0.8076 untuk PCA-SVM dan 0.7978 untuk SVM

#### REFERENSI

- [1] Badan Pusat Statistik, "Curah Hujan di Stasiun Pengamatan Klimatologi Bogor Menurut Bulan (mm), 2020-2022," Badan Pusat Statistik, 2022. [Online]. Available: <https://jabar.bps.go.id/indicator/151/430/1/-curah-hujan-di-stasiun-pengamatan-klimatologi-bogor-menurut-bulan.html>. [Accessed September 2023].
- [2] PAM JAYA, "Sejarah PAM JAYA," PAM JAYA, 2023. [Online]. Available: <https://www.pamjaya.co.id/tentangkami>. [Accessed September 2023].
- [3] Open Data Jabar, "Kabupaten Bogor Berpenduduk Terbanyak Se-Indonesia," Open Data Jabar, 23 Agustus 2020. [Online]. Available: <https://opendata.jabarprov.go.id/id/infografik/kabupaten-bogor-berpenduduk-terbanyak-se-indonesia>. [Accessed September 2023].
- [4] A. Ikhsan, "8.900 KK di Kabupaten Bogor Alami Krisis Air Bersih karena Kekeringan," Kompas, 4 August 2023. [Online]. Available: <https://bandung.kompas.com/read/2023/08/04/161555078/8900-kk-di-kabupaten-bogor-alami-krisis-air-bersih-karena-kekeringan?page=all>. [Accessed September 2023].
- [5] Angellina, D. E. Herwindiati and J. Hendryli, "Performa Support Vector Machine Pada Klasifikasi," *Jurnal Media Informatika Budidarma*, vol. 7, pp. 231-241, 2023.
- [6] T. Widodo, "Kajian Ketersediaan Air Tanah Terkait

Pemanfaatan Lahan di Kabupaten Blitar," *Jurnal Pembangunan Wilayah dan Kota*, vol. 9, no. 2, pp. 122-133, 2013.

- [7] D. Budi, D. E. Herwindiati and J. Hendryli, "Land Use Change Using Least Absolute Shrinkage and Selection Operator," *Symposium on Computer Applications & Industrial Electronics*, vol. 11, no. 1, pp. 30-35, 2021.
- [8] L. I. Smith, "A Tutorial on Principal Components Analysis," *Semantic Scholar*, 2002.
- [9] M. M. Adankon and M. Cheriet, "Support Vector Machine," *Encyclopedia of Biometrics*, p. 1303-1308, 2009.

**Stephen Adhikurnia**, mahasiswa tingkat akhir Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jakarta.

**Dyah Erny Herwindiati**, dosen Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jakarta.

**Janson Hendryli**, dosen Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jakarta.