

SKRIPSI

**PENGUMPULAN DATA CITRA SATELIT KELURAHAN
DAN PERHITUNGAN LUAS AREA HIJAU DENGAN
TEKNOLOGI BIG DATA**



Juan Anthonius Kusjadi

NPM: 6181801059

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2022**

UNDERGRADUATE THESIS

**COLLECTING SATELLITE IMAGES OF SUBDISTRICTS
AND COMPUTING GREEN AREAS USING BIG DATA
TECHNOLOGY**



Juan Anthonius Kusjadi

NPM: 6181801059

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2022**

LEMBAR PENGESAHAN

PENGUMPULAN DATA CITRA SATELIT KELURAHAN DAN PERHITUNGAN LUAS AREA HIJAU DENGAN TEKNOLOGI BIG DATA

Juan Anthonius Kusjadi

NPM: 6181801059

Bandung, 02 Agustus 2022

Menyetujui,

Pembimbing

Digitally signed
by Veronica Sri
Moertini

Dr. Veronica Sri Moertini

Ketua Tim Penguji

Digitally signed
by Luciana
Abednego

Luciana Abednego, M.T.

Anggota Tim Penguji

Digitally signed
by Lionov

Lionov, Ph.D.

Mengetahui,

Ketua Program Studi

Digitally signed
by Mariskha Tri
Adithia

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

PENGUMPULAN DATA CITRA SATELIT KELURAHAN DAN PERHITUNGAN LUAS AREA HIJAU DENGAN TEKNOLOGI BIG DATA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 02 Agustus 2022



Juan Anthonius Kusjadi
NPM: 6181801059

ABSTRAK

Pembangunan dan perkembangan wilayah cenderung mengakibatkan pengalihfungsian Ruang Terbuka Hijau (RTH) dan membuat luasnya semakin sedikit. Menurunnya luas RTH dapat menurunkan jumlah oksigen yang dihasilkan wilayah tersebut. Di Indonesia luas minimum RTH pada wilayah perkotaan dan kelurahan sudah ada undang-undang yang mengaturnya. Melanjutkan dari penelitian sebelumnya yang melakukan analisis pada wilayah kelurahan-kelurahan di Kota Bandung, penelitian ini melakukan analisis pada wilayah kelurahan-kelurahan di beberapa Kota di Indonesia.

Analisis pada citra satelit dapat digunakan untuk mengetahui luas RTH pada suatu wilayah. Perhitungan luas RTH dari citra satelit merupakan pendekatan yang dilakukan untuk mengestimasi luas wilayah sesungguhnya. Citra satelit yang digunakan adalah citra satelit kenampakan bumi dari atas. Citra satelit dapat diambil dari *Tile Map Server* yang dimiliki oleh ArcGIS. Tingkat kedetailan gambar citra satelit akan dipengaruhi oleh tingkat perbesaran citra satelit yang diambil.

Meningkatnya data citra satelit membuat kebutuhan komputasi menjadi meningkat juga. Teknologi big data dapat digunakan untuk mengimbangi meningkatnya data citra satelit. *Framework* teknologi big data yang sering digunakan adalah Apache Hadoop dan Apache Spark. Kedua *framework* tersebut mengimplementasikan komputasi yang terdistribusi. Apache Hadoop dapat digunakan dalam pengambilan data citra satelit menggunakan skema *MapReduce* dan disimpan pada *Hadoop Distributed File System* (HDFS). Apache Spark dapat digunakan untuk melakukan perhitungan luas pada citra satelit dengan memanfaatkan *Spark Machine Learning Library* (MLlib). Segmentasi pada citra dilakukan dengan menggunakan model algoritma klasterisasi *KMeans*. Hasil segmentasi digunakan untuk menghitung luas wilayah kelurahan dan wilayah hijau.

Proses mengunduh citra satelit dan memotong sesuai batas wilayah dilakukan satu persatu pada masing-masing *tile* kelurahan. Hal tersebut dilakukan untuk mengurangi penggunaan memori yang membesar ketika memuat gambar menjadi satu baru dipotong dikarenakan ada kelurahan yang sangat besar. Segmentasi dilakukan pada fitur warna citra satelit. Fitur warna citra satelit memiliki ukuran yang jauh lebih besar dari ukuran berkas citra satelit. Hasil segmentasi dipakai untuk menghitung luas RTH yang ditumbuhi pepohonan dan luas kelurahan. Perhitungan luas kelurahan menggunakan metode *pixel based*. Evaluasi dilakukan pada perhitungan luas kelurahan dengan luas kelurahan sebenarnya. Hasil segmentasi dan perhitungan luas ditampilkan pada sebuah perangkat lunak.

Perangkat lunak untuk menampilkan hasil analisis dan segmentasi dibangun menggunakan *Framework Java Spring*. Arsitektur yang digunakan dalam membentuk perangkat lunak adalah arsitektur *Model, View, Controller* (MVC). *Framework Bootstrap* digunakan untuk membantu membentuk tampilan pada perangkat lunak agar lebih menarik. Salah satu hasil yang didapatkan adalah presentase RTH pada Kota Bandung yaitu sebesar 23.85% dari keseluruhan luas Kota Bandung.

Kata-kata kunci: Sistem Terdistribusi, Komputasi Pararel, Apache Spark, Apache Hadoop, Machine Learning, Hadoop FileSystem, MapReduce, K-Means, Scala, Klasterisasi, Citra Satelit

ABSTRACT

Regional development tends to convert Green Open Space and decrease its area. The decrease in the green open space area can reduce the amount of oxygen produced. In Indonesia, the minimum area of green open space in urban areas has been regulated. Continuing from previous research that analyzed urban villages in Bandung City, this study analyzed subdistrict in several cities in Indonesia.

Analysis of satellite imagery can be used to determine the area of green open space in an area. Calculation of the area of green open space from satellite imagery is an approach taken to estimate the actual area. One of the satellite images that can be used is satellite imagery of the earth's appearance from above. Satellite images can be retrieved from the Tile Map-Server owned by ArcGIS. The level of detail of satellite imagery will be influenced by the level of magnification of the satellite image taken.

The increase in satellite image data affected the need for computing increase. Big data technology can be used to compensate for increasing satellite imagery data. The Big data technology frameworks that are often used are Apache Hadoop and Apache Spark. Both of these frameworks implement distributed computing. Apache Hadoop can be used to retrieve satellite image data using the MapReduce scheme and stored in the Hadoop Distributed File System (HDFS). Apache Spark can be used to perform extensive calculations on satellite imagery by utilizing the Spark Machine Learning Library (MLlib). Image segmentation is done using the KMeans clustering algorithm model. The segmentation results are used to calculate the area of subdistrict and green areas.

The process of downloading and cutting satellite images based on regional boundaries data is carried out one by one for each subdistrict satellite image tile. This is done to reduce the use of RAM that enlarges when loading images into one and then cut it because there is a very large area to process. Segmentation is performed on the color features of the satellite image. The color feature of the satellite image has a much larger size than the file size of the satellite image. The results of the segmentation are used to calculate the area of green open space-grown with trees and the area of the village. Calculation of the area of the village using the pixel-based method. The evaluation is carried out on the calculation of the area of the subdistrict with the actual area of the subdistrict. The results of segmentation and area calculations are displayed in the software.

The software for displaying analysis and segmentation results is built using the Java Spring Framework. The architecture used in forming the software is the Model, View, Controller (MVC) architecture. The Bootstrap framework is used to help shape the appearance of the software to make it more attractive. One of the results obtained is the percentage of green open space in the Bandung City, which is 23.85% from its total area.

Keywords: Distributed System, Parallel Computation, Apache Spark, Apache Hadoop, Machine Learning, Hadoop FileSystem, MapReduce, K-Means, Scala, Clustering, Satellite Imagery

*Kepada keluarga dan masa depan Ruang Terbuka Hijau di
Indonesia*

KATA PENGANTAR

Puji syukur kepada Tuhan yang Maha Esa karena atas berkat-Nya penulis dapat menyelesaikan dokumen penelitian skripsi ini dengan baik dan tepat waktu. Skripsi ini dibuat sebagai salah satu syarat kelulusan untuk menyelesaikan pendidikan Strata 1 (S1) pada Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Sains (FTIS), Universitas Katolik Parahyangan. Dalam prosesnya menyusun skripsi ini tentu tidaklah mudah dan ada berbagai tantangan dan rintangan yang sudah dilalui. Oleh karena itu, berkat dukungan dan bantuan yang diberikan oleh berbagai pihak, penulis ingin mengucapkan terima kasih yang sebesar-besarnya diberikan kepada:

1. Orang tua penulis, yang selalu mendoakan dan memberikan dukungan baik materil dan immateril pada penulis dalam menyelesaikan skripsi ini.
2. Ibu Dr. Veronica Sri Moertini selaku dosen pembimbing yang telah bersedia memberikan waktu, ilmu, kritik, saran dan dengan sabar membimbing dan mengarahkan penulis agar dapat menyelesaikan skripsi ini dengan baik dan tepat waktu.
3. Teman-teman yang tidak dapat disebutkan satu persatu yang sudah memberikan dukungan dalam menyelesaikan skripsi ini.

Bandung, Agustus 2022

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxiii
DAFTAR KODE PROGRAM	xxvi
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Batasan Masalah	4
1.5 Metodologi	5
1.6 Sistematika Pembahasan	5
2 LANDASAN TEORI	7
2.1 Ruang Terbuka Hijau	7
2.2 RTH di Wilayah Kelurahan	8
2.3 Citra Satelit	8
2.4 Teknologi Big Data	11
2.4.1 <i>Apache Hadoop</i>	13
2.4.2 <i>Apache Spark</i>	15
2.5 <i>Algoritma Klasterisasi</i>	16
2.5.1 <i>K-Means</i>	16
2.5.2 <i>Metode Elbow</i>	17
2.6 Algoritma K-Means Paralel	17
2.7 <i>Data Lake</i>	20
2.8 <i>Base64</i>	22
2.9 <i>JSON</i>	22
3 PENGUMPULAN, PENYIAPAN, DAN EKSPLORASI DATA	23
3.1 Pengumpulan Data Citra Satelit	23
3.1.1 Mengunduh Data Geospasial	23
3.1.2 Pembuatan dan Penyiapan Data Input Hadoop MapReduce	25
3.1.3 Mengunduh dan Menyiapkan Data Citra Satelit menggunakan Hadoop Map-Reduce	27
3.2 Membuat Model Klasterisasi Data Citra Satelit	37
3.3 Memproses Data Citra Satelit	39
3.3.1 Penyiapan Data citra satelit	40
3.3.2 Menjalankan model klasterisasi K-Means	40

4	ANALISIS HASIL PENGUMPULAN DAN SEGMENTASI CITRA SATELIT	45
4.1	Analisis Waktu untuk Mengumpulkan Data Citra Satelit	45
4.2	Analisis Hasil Pengumpulan Data Citra Satelit	46
4.3	Analisis Waktu untuk Memproses Data Citra Satelit	48
4.4	Perbandingan Waktu untuk Mengunduh dan Memproses Data Citra Satelit	49
4.5	Evaluasi Perhitungan Luas Wilayah Kelurahan	49
4.6	Evaluasi Perhitungan Luas Wilayah Kota	50
5	PEMBANGUNAN <i>Data Lake</i> DAN PERANGKAT LUNAK	53
5.1	Arsitektur Sistem Data Lake	53
5.2	Struktur data pada <i>Data Lake</i>	55
5.2.1	Struktur Berkas Data <i>Tile</i> Citra Satelit	55
5.2.2	Struktur Berkas Data <i>tile</i> Citra Satelit yang Sudah Dipotong	56
5.2.3	Struktur Berkas Data Metadata dan Geospasial	57
5.2.4	Struktur Berkas Data Model Klasterisasi	58
5.2.5	Struktur Berkas Data Hasil Klasterisasi	58
5.3	Arsitektur Perangkat Lunak	59
5.4	Fitur-Fitur Perangkat Lunak	60
5.5	Diagram Use Case Perangkat Lunak	60
5.6	Perancangan Antar Muka Perangkat Lunak	63
5.6.1	Halaman Home	63
5.6.2	Halaman Explorer	64
5.6.3	Halaman Download	64
5.6.4	Halaman Process	65
5.6.5	Halaman Result	65
5.6.6	Halaman Preview Result	66
5.7	Diagram Kelas Perangkat Lunak	66
5.7.1	Kelas HomeController	66
5.7.2	Kelas ExplorerController	66
5.7.3	Kelas ResultExplorerController	67
5.7.4	Kelas ResultData	67
5.7.5	Kelas StreamRunnable	68
5.7.6	Kelas DownloadController	68
5.7.7	Kelas ResultController	68
5.8	Implementasi Perangkat Lunak	68
5.8.1	Halaman <i>Home</i>	68
5.8.2	Halaman <i>Download</i>	69
5.8.3	Halaman <i>Process</i>	70
5.8.4	Halaman Explorer	70
5.8.5	Halaman Result	71
5.8.6	Halaman Preview Result	71
6	KESIMPULAN DAN SARAN	73
6.1	Kesimpulan Penelitian	73
6.2	Saran Pengembangan	73
	DAFTAR REFERENSI	75
	A KODE PROGRAM	77
	B HASIL EKSPERIMEN	119

DAFTAR GAMBAR

1.2	Gambar citra satelit dari ArcGIS di Kelurahan Ciateul, Kota Bandung, Jawa Barat	2
1.3	Ilustrasi hasil pemotongan setiap tile di Kelurahan Ciateul, Kota Bandung, Jawa Barat	3
2.1	Citra satelit dari ArcGIS di Kelurahan Ciateul, Kota Bandung, Jawa Barat ¹	8
2.2	Penggambaran posisi garis lintang (<i>latitude</i>) dan garis bujur (<i>longitude</i>) ²	9
2.3	Ilustrasi citra satelit yang diproyeksikan dari bentuk bulat menjadi silinder dengan menggunakan proyeksi mercator ³	9
2.4	Ilustrasi pembagian <i>tile</i> pada tingkat perbesaran satu dan dua. Ketika tingkat perbesaran bertambah menyebabkan sebuah <i>tile</i> pada tingkat perbesaran sebelumnya terbagi menjadi empat buah <i>tile</i> .	10
2.5	Detail karakteristik dari Big Data [1]	11
2.6	Penggambaran dari perumpamaan fenomena <i>data deluge Big Data</i> ⁴	12
2.7	Arsitektur <i>master-slave</i> pada <i>hadoop</i> [2].	13
2.8	Penyerahan pekerjaan dari client ke <i>MapReduce</i> , pekerjaan diurai menjadi map kecil dan beberapa <i>reduce task</i> [2].	14
2.9	Arsitektur dari aplikasi <i>Spark</i> untuk sebuah klaster. <i>Spark Session</i> (kiri) memiliki akses untuk mengirimkan pecahan program dari <i>user</i> ke masing-masing <i>executor</i> (kanan) [3].	15
2.10	Contoh hasil klasterisasi dari objek yang berupa titik di bidang 2 dimensi	18
2.11	Metode <i>elbow</i> . Sumbu <i>x</i> pada gambar merupakan jumlah <i>k</i> dan sumbu <i>y</i> merupakan nilai <i>cost</i> ketika algoritma dijalankan pada <i>k</i> . Dinamakan metode <i>elbow</i> dikarenakan <i>k</i> yang dipilih sebagai hasil ada pada siku dari grafik seperti pada lingkaran merah.	18
2.12	Diagram alur cara kerja K-Means pada Spark MLlib [4]	20
2.13	Algoritma k-Means paralel pada Spark dengan kasus jumlah kelompok = 2 ($mP = \text{mapPartition}$, $rBK = \text{reduceByKey}$) [5]	21
3.1	Ilustrasi proses penyiapan data <i>input</i> untuk Hadoop MapReduce. Bentuk data pada ilustrasi merupakan gambaran dari data sebenarnya	23
3.2	Potongan Data atribut dan koordinat dalam format <i>JSON</i> . Data yang digunakan adalah seluruh data yang berisi pada <i>key features</i> .	24
3.3	Data geospasial yang disimpan pada Hadoop HDFS dengan format CSV	25
3.4	Koordinat batas ujung-ujung dari Kelurahan Ciateul, Kota Bandung, Jawa Barat	26
3.5	Ilustrasi posisi <i>tile</i> yang didapatkan setelah perhitungan konversi dari Titik T1, T2, T3, dan T4 yang ditandai pada kotak berwarna merah muda.	27
3.6	Ilustrasi untuk menggambarkan Map-Side Join Pattern. Pada bagian kiri adalah ilustrasi dari data <i>input</i> Mapper. Pada bagian kanan adalah HashTable dari data <i>tile</i> yang sudah pernah diunduh yang didapatkan dari <i>distributed cache files</i>	28
3.7	Ilustrasi untuk menunjukkan sebuah <i>tile</i> dapat digunakan oleh lebih dari 1 kelurahan. Sebuah kelurahan ditandai dengan warna. Jika pada sebuah <i>tile</i> terdapat lebih dari 1 warna maka <i>tile</i> tersebut digunakan oleh lebih dari 1 kelurahan	30
3.8	Contoh bentuk data <i>tile</i> citra satelit yang disimpan pada <i>data lake</i>	30
3.9	Contoh bentuk data <i>output</i> Mapper untuk Reducer	31

3.10	Ilustrasi konversi koordinat <i>degree</i> menjadi koordinat piksel	32
3.11	Ilustrasi penggeseran poligon	33
3.12	Ilustrasi Pemotongan dengan metode Intersect antara Area kelurahan dan Area <i>Tile</i>	34
3.13	Ilustrasi Penggeseran Area <i>Tile</i> yang sudah dipotong ke posisi 0,0	35
3.14	Ilustrasi proses pemotongan <i>tile</i> citra satelit	35
3.15	Ilustrasi perhitungan luas sebuah pixel pada koordinat <i>tile</i>	36
3.16	Ilustrasi contoh data <i>output</i> Hadoop MapReduce	37
3.17	Ilustrasi pembuatan model klasterisasi K-Means menggunakan Spark MLlib	37
3.18	Data <i>input</i> yang digunakan untuk membangun model klasterisasi	37
3.19	Proses membongkar data gambar dan mengambil fitur warna setiap piksel	38
3.20	Data <i>input</i> untuk pembangunan model	38
3.21	Data akhir fitur untuk pembangunan model	39
3.22	Hasil fungsi evaluasi model klasterisasi K-Means untuk setiap k dari 2 sampai 15	39
3.23	Ilustrasi klasterisasi K-Means menggunakan Spark MLlib	40
3.24	Data yang sudah disiapkan dengan fitur warna	40
3.25	Data yang sudah disiapkan untuk input model klasterisasi K-Means	41
3.26	Data hasil prediksi klasterisasi K-Means	41
3.27	Data hasil prediksi yang sudah digabungkan menjadi satu untuk setiap <i>tile</i>	41
3.28	Data hasil perhitungan luas	42
3.29	Seluruh citra satelit sebelum dan sesudah disegmentasi menggunakan algoritma K-Means dengan k=5 dan disusun sesuai dengan posisi x dan y koordinat <i>tile</i> pada Kelurahan Sukamaju	42
3.30	Seluruh citra satelit sebelum dan sesudah disegmentasi menggunakan algoritma K-Means dengan k=5 dan disusun sesuai dengan posisi x dan y koordinat <i>tile</i> pada Kelurahan Ciateul	43
3.31	Seluruh citra satelit sebelum dan sesudah disegmentasi menggunakan algoritma K-Means dengan k=5 dan disusun sesuai dengan posisi x dan y koordinat <i>tile</i> pada Kelurahan Ciumbuleuit	43
3.32	Seluruh citra satelit sebelum dan sesudah disegmentasi menggunakan algoritma K-Means dengan k=5 dan disusun sesuai dengan posisi x dan y koordinat <i>tile</i> pada Kelurahan Batununggal	44
4.1	Hasil citra berasal dari Kota Bogor	46
4.2	Hasil citra berasal dari Kabupaten Cirebon	47
4.3	Hasil citra berasal dari Kota Bandung	47
5.1	Arsitektur sistem <i>data lake</i>	53
5.2	Struktur folder <i>data lake</i>	54
5.3	Struktur berkas data <i>tile</i> citra satelit	55
5.4	Struktur data <i>tile</i> citra satelit	55
5.5	Struktur berkas data <i>tile</i> citra satelit yang sudah dipotong	56
5.6	Struktur data <i>tile</i> citra satelit yang sudah dipotong	56
5.7	Struktur berkas data metadata dan geospatial	57
5.8	Struktur data meta dan geospatial	57
5.9	Struktur berkas data model klasterisasi	58
5.10	Struktur berkas data hasil klasterisasi	58
5.11	Struktur data hasil klasterisasi	59
5.12	Ilustrasi hubungan <i>data lake</i> dengan perangkat lunak	59
5.13	Diagram <i>Use Case</i>	61
5.14	Rancangan halaman <i>home</i>	63
5.15	Rancangan halaman <i>explorer</i>	64
5.16	Rancangan halaman <i>download</i>	64

5.17 Rancangan halaman <i>process</i>	65
5.18 Rancangan halaman <i>result</i>	65
5.19 Diagram rancangan halaman <i>preview result</i>	66
5.20 Diagram kelas	67
5.21 Halaman <i>Home</i>	69
5.22 Halaman <i>Download</i>	69
5.23 Halaman <i>Process</i>	70
5.24 Halaman <i>Explorer</i>	70
5.25 Halaman <i>Result</i>	71
5.26 Halaman Preview Result	71

DAFTAR TABEL

2.1	Parameter yang digunakan pada kelas yang mengimplementasi algoritma K-Means pada Spark	19
2.2	Karakteristik Data Lake	21
3.1	Parameter kembalian yang digunakan pada API	24
3.2	Kolom data kembalian yang diminta dari API	24
3.3	Parameter yang dibutuhkan ketika mengonversi dari koordinat <i>degree</i> ke koordinat <i>tile</i>	26
3.4	Contoh Data <i>input</i> Mapper untuk <i>pos_x</i> dengan nilai 52358	27
3.5	Parameter yang dibutuhkan ketika mengonversi dari koordinat <i>degree</i> ke koordinat <i>piksel</i>	32
3.6	Nilai konversi dari satuan DMS ke kilometer ⁵	35
4.1	Perhitungan waktu yang diperlukan untuk mengumpulkan data citra satelit pada beberapa wilayah berbeda	45
4.2	Spesifikasi klaster Hadoop diambil menggunakan perintah " <i>free -m</i> " dan " <i>lsproc</i> " namun pada YARN alokasi maksimum hanya 8 GigaByte	46
4.3	Perhitungan waktu yang diperlukan untuk memproses data citra satelit pada beberapa wilayah berbeda. (Tanda strip "-" menandakan tidak ada data dikarenakan kegagalan saat eksekusi)	48
4.4	Perhitungan total waktu mengunduh dan memproses data citra satelit antara yang menggunakan komputasi terdistribusi dan tidak	49
4.5	Hasil perhitungan luas menggunakan k=5 dengan metode <i>pixel based</i>	50
4.6	Hasil perhitungan luas minimal RTH kelurahan	50
4.7	Hasil perhitungan luas minimal RTH perkotaan	51
5.1	Tabel kebutuhan halaman dan fitur	60
B.1	Hasil perhitungan luas wilayah kelurahan dan wilayah RTH Kelurahan di Kota Bandung menggunakan algoritma KMeans dengan k=5 dan pendekatan <i>pixel based</i>	119
B.2	Hasil perhitungan luas wilayah kelurahan dan wilayah RTH Kelurahan di Kota Bogor menggunakan algoritma KMeans dengan k=5 dan pendekatan <i>pixel based</i>	123
B.3	Hasil perhitungan luas wilayah kelurahan dan wilayah RTH Kelurahan di Kota Cirebon menggunakan algoritma KMeans dengan k=5 dan pendekatan <i>pixel based</i>	124

DAFTAR KODE PROGRAM

2.1	Contoh notasi JSON yang sah	22
A.1	Tile.java	77
A.2	Coordinate.java	77
A.3	GeoData.java	78
A.4	Pair.java	78
A.5	IntermediateTileData.java	79
A.6	ProvinceCity.java	79
A.7	CommonUtils.java	80
A.8	Base64Utils.java	80
A.9	GeoDataUtils.java	81
A.10	GeoSpatialUtils.java	83
A.11	HttpUtils.java	84
A.12	LoggerUtils.java	84
A.13	StopWatchUtils.java	85
A.14	TileMapUtils.java	85
A.15	TileProcessorUtils.java	86
A.16	GeoDataDownloaderThreads.java	87
A.17	GeoDataProcessorThreads.java	88
A.18	GeoDataDownloaderJob.java	88
A.19	StoragePreparationJob.java	90
A.20	MapDownloaderJob.java	91
A.21	PostProcessJob.java	92
A.22	ReportingJob.java	94
A.23	MapDownloaderMapper.java	95
A.24	MapDownloaderReducer.java	97
A.25	ImageUtils.scala	99
A.26	ImageProcessor.scala	101
A.27	KMeansTrainer.scala	105
A.28	ResultData.java	106
A.29	StreamRunnable.java	106
A.30	HomeController.java	106
A.31	ExplorerController.java	107
A.32	DownloadController.java	108
A.33	ProcessController.java	109
A.34	ResultController.java	110
A.35	defaultView.html	112
A.36	header.html	112
A.37	metahead.html	113
A.38	index.html	113
A.39	explorer.html	114
A.40	downloader.html	115

A.41 process.html	115
A.42 resultPreview.html	116
A.43 explorer.html	117

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pembangunan dan perkembangan suatu wilayah cenderung mengakibatkan pengalihfungsian Ruang Terbuka Hijau (RTH) untuk memenuhi kebutuhan tempat tinggal bagi masyarakat kota. Masyarakat kota memerlukan oksigen untuk hidup dan beraktivitas (lihat Gambar 1.1). Menurut Undang-Undang (UU) Nomor 26 Tahun 2007 tentang Penataan Ruang Pasal 1 Ayat 31¹, ruang Terbuka hijau adalah area memanjang/jalur dan/atau mengelompok, yang penggunaannya lebih bersifat terbuka, tempat tumbuh tanaman, baik yang tumbuh secara alamiah maupun yang sengaja ditanam. RTH merupakan salah satu penyumbang oksigen bagi kota. RTH seharusnya tersedia dalam jumlah yang cukup, terutama pada perkotaan yang padat penduduk. Menurut pasal 29 ayat 2, Undang-undang nomor 26 tahun 2007 tentang Penataan Ruang menyebutkan bahwa proporsi ruang terbuka hijau pada wilayah kota paling sedikit 30% dari luas wilayah kota.



Gambar 1.1: Ilustrasi Ruang Terbuka Hijau²

Salah satu cara untuk dapat mengetahui luas RTH pada suatu kota adalah dengan memanfaatkan citra satelit. Citra satelit merupakan gambaran kenampakan bumi yang didapatkan dari satelit (lihat Gambar 1.2). Citra satelit saat ini mudah untuk didapatkan. ArcGIS³ merupakan salah satu aplikasi yang menyediakan citra satelit secara gratis. Citra satelit dapat digunakan untuk mengidentifikasi kelompok pepohonan pada suatu wilayah dan dihitung luasnya. Hasil perhitungan luas RTH pada suatu wilayah diharapkan dapat memberikan dorongan untuk meningkatkan pengijauan dan/atau digunakan oleh pemerintah untuk merancang dan meningkatkan penghijauan di berbagai wilayah di Indonesia.

¹Undang-Undang Nomor 26 Tahun 2007 tentang Penataan Ruang. <http://dcpk3.bandung.go.id/storage/app/uploads/public/580/eb8/182/580eb8182fd68859480338.pdf>

²Ilustrasi ruang terbuka hijau <https://asiatoday.id/read/who-ruang-terbuka-hijau-selamatkan-warga-kota-dari-kematian-dini>

³Situs penyedia data citra satelit ArcGIS <https://www.arcgis.com/>



Gambar 1.2: Gambar citra satelit dari ArcGIS di Kelurahan Ciateul, Kota Bandung, Jawa Barat

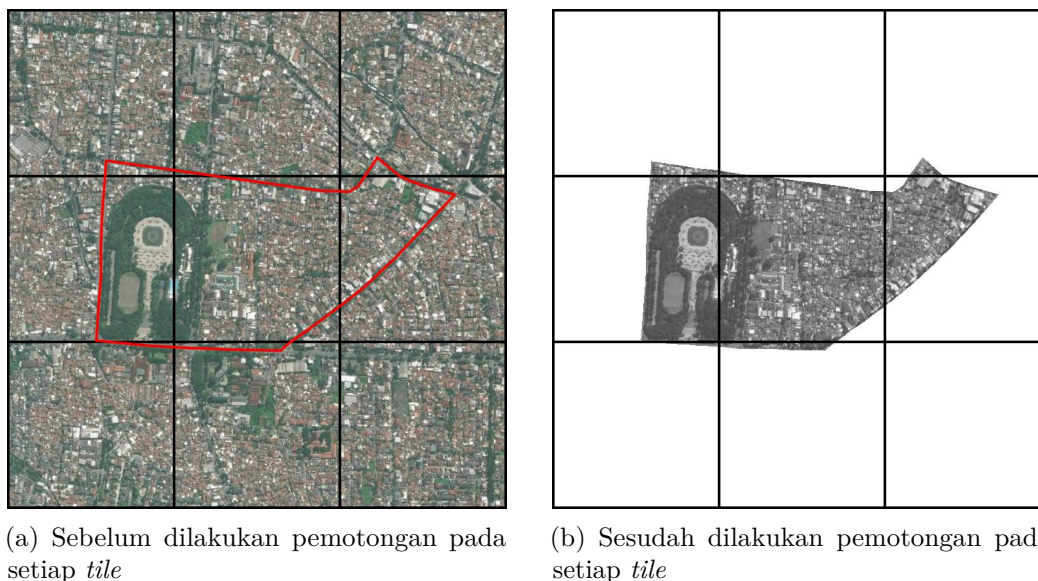
Penelitian sejenis untuk menghitung luas RTH sudah dilakukan dengan metode *image segmentation* pada citra satelit sudah dilakukan oleh Fritz Humphrey Silalahi [6]. Pada penelitian tersebut sudah dilakukan eksperimen terhadap metode dan jumlah k yang baik untuk memproses data citra satelit menggunakan algoritma K-Means. Penelitian tersebut sudah dilakukan untuk beberapa kelurahan di Kota Bandung. Dari penelitian tersebut dapat ditarik kesimpulan bahwa segmentasi citra menggunakan algoritma K-Means sudah menghasilkan segmentasi yang baik. Segmentasi citra tersebut dihasilkan dari penggunaan metode *pixel based* dan k kluster dengan nilai 5.

Pada penelitian yang dilakukan oleh Fritz Humphrey Silalahi [6], proses pengumpulan batas wilayah masih dilakukan secara manual. Hal tersebut menyebabkan pemrosesan menjadi terbatas pada data batas wilayah yang dimiliki. Pada penelitian tersebut juga pengunduhan terhadap *tile* citra satelit masih dilakukan satu-satu sehingga banyak *tile* citra satelit yang sama diunduh beberapa kali. Algoritma klusterisasi K-Means yang digunakan masih bersifat non-paralel sehingga kemampuan komputasi terbatas pada kemampuan satu komputer saja. Pada penelitian tersebut juga pemrosesan hanya dilakukan pada kelurahan-kelurahan di Kota Bandung.

Solusi yang ingin ditawarkan pada penelitian ini adalah melakukan otomatisasi pengumpulan data citra satelit dan mencoba memperluas jangkauan perhitungan kelurahan-kelurahan wilayah RTH dengan menggunakan algoritma K-Means paralel. Solusi tersebut diharapkan dapat memperbaiki kekurangan yang terdapat pada penelitian Fritz Humphrey Silalahi [6]. Solusi untuk penelitian tersebut dicoba dilakukan dengan bantuan teknologi *big data*. Teknologi *big data* menawarkan solusi untuk pengolahan data citra satelit yang sangat besar. Teknologi *big data* yang digunakan pada penelitian ini adalah Apache Hadoop dan Apache Spark.

Apache Hadoop merupakan sebuah *framework* yang menyediakan sistem penyimpanan menggunakan *Hadoop Distributed File System* (HDFS) dan komputasi secara terdistribusi menggunakan Hadoop MapReduce. Kedua fungsi tersebut dipisahkan sejak hadoop versi 3. Apache Spark menyediakan berbagai *tools* untuk mengakomodasi komputasi terdistribusi. Apache Spark dapat menjadi *tools* untuk menutupi kekurangan pada Hadoop MapReduce yaitu seperti melakukan *query SQL* (*Structured Query Language*) dan menggunakan algoritma *machine learning* pada Spark MLlib [7]. Berbeda dengan Apache Hadoop yang secara dasar mendukung bahasa pemrograman Java, Apache Spark mendukung beberapa bahasa pemrograman yaitu Java, Python, R dan Scala.

Komputasi terdistribusi merupakan teknik komputasi dengan membagi data yang besar menjadi pecahan-pecahan untuk diproses menjadi kumpulan tugas-tugas (*task*) kecil di beberapa komputer secara paralel [8]. Komputasi terdistribusi memungkinkan pengolahan data dengan lebih mudah bagi algoritma sehingga mendapatkan hasil yang lebih akurat dengan waktu yang lebih singkat [8].



Gambar 1.3: Ilustrasi hasil pemotongan setiap *tile* di Kelurahan Ciuteul, Kota Bandung, Jawa Barat

Komputasi terdistribusi membutuhkan beberapa komputer yang terkoneksi dalam sebuah jaringan. Pada umumnya komputasi terdistribusi memiliki arsitektur *master* dan *slave* dimana sebuah *master* memiliki beberapa *slave*.

Pada penelitian ini citra satelit dari sebuah provinsi di Indonesia dipecah menjadi beberapa kelurahan untuk setiap kota. Data citra satelit di setiap kota dipecah lagi menjadi beberapa *tile*. *Tile* merupakan sistem indeks yang digunakan untuk menandai suatu wilayah dalam *grid* citra satelit. Setiap *tile* gambar dipotong sesuai dengan data *polygon* (lihat Gambar 1.3) dari web Badan Informasi Geospasial Indonesia⁴. Proses penyiapan data citra satelit ini menggunakan Hadoop MapReduce. Data citra kelurahan kemudian disimpan pada Hadoop HDFS.

Data citra disegmentasi menggunakan algoritma klasterisasi K-Means pada Apache Spark untuk setiap *tile* pada kelurahan. Luas wilayah RTH yang ditumbuhi pepohonan dihitung berdasarkan hasil klasterisasi dengan pendekatan *pixel based*. Pendekatan tersebut menghitung luas dengan menjumlahkan piksel-piksel yang teridentifikasi sebagai area RTH. Pendekatan tersebut menghasilkan jumlah piksel yang teridentifikasi area RTH. Luas wilayah RTH dihitung dengan mengkalikan jumlah piksel suatu wilayah dengan ukuran piksel per satuan luas. Satuan luas yang digunakan adalah kilometer persegi (km^2). Luas RTH kelurahan dihitung dengan menjumlahkan luas RTH setiap *tile* pada sebuah kelurahan.

Hasil perhitungan luas RTH kelurahan yang ditumbuhi pepohonan dibandingkan dengan batas minimum RTH kelurahan pada Peraturan Menteri Pekerjaan Umum Nomor 05/PRT/M/2008 tentang Pedoman Penyediaan dan Pemanfaatan Ruang Terbuka Hijau di Kawasan Perkotaan⁵. Hasil perhitungan luas juga dibandingkan dengan luas sesungguhnya berdasarkan data dari Badan Informasi Geospasial Indonesia. untuk mengukur besar akurasi. Hasil dari analisis luas wilayah RTH pada kelurahan ditampilkan pada sebuah perangkat lunak berbasis web. Hasil analisis bertujuan untuk dapat memberikan informasi kepada masyarakat untuk lebih peduli dengan lingkungan hidup sekitarnya. Hasil analisis juga dapat menjadi pedoman untuk pemerintah agar dapat membuat regulasi terkait perluasan wilayah RTH.

⁴Situs penyedia data Geospasial Indonesia <https://geoservices.big.go.id>

⁵Peraturan Menteri Pekerjaan Umum Nomor 05/PRT/M/2008 tentang Pedoman Penyediaan dan Pemanfaatan Ruang Terbuka Hijau di Kawasan Perkotaan <http://sim.ciptakarya.pu.go.id/p2kh/knowledge/detail/permen-pu-05-2008-rth>

1.2 Rumusan Masalah

Rumusan masalah yang muncul berdasarkan deskripsi dan latar belakang yang sudah dibahas adalah sebagai berikut:

1. Bagaimana mengumpulkan data berupa citra satelit dari beberapa kelurahan dan kota pada sebuah provinsi di Indonesia?
2. Bagaimana menyiapkan data berupa citra satelit dari beberapa kelurahan dan kota pada sebuah provinsi di Indonesia?
3. Bagaimana mendeteksi dan menghitung luas RTH dari beberapa kelurahan dan kota pada sebuah provinsi di Indonesia?
4. Bagaimana mengevaluasi hasil deteksi dan perhitungan RTH dari beberapa kelurahan dan kota pada sebuah provinsi di Indonesia?
5. Bagaimana melakukan otomatisasi pengumpulan dan pemrosesan data RTH dari beberapa kelurahan dan kota pada sebuah provinsi di Indonesia?
6. Bagaimana menampilkan hasil segmentasi dari beberapa kelurahan dan kota pada sebuah provinsi di Indonesia pada perangkat lunak?

1.3 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengumpulkan data berupa citra satelit dari kelurahan pada beberapa kota di Indonesia pada Hadoop HDFS.
2. Menyiapkan data berupa citra satelit dari kelurahan pada beberapa kota di Indonesia.
3. Menganalisis dan menghitung luas dari RTH dari kelurahan pada beberapa kota di Indonesia menggunakan Apache Spark.
4. Mengevaluasi hasil analisis dan perhitungan luas RTH dari kelurahan pada beberapa kota di Indonesia.
5. Melakukan otomatisasi dalam pengumpulan dan pemrosesan data citra satelit dari kelurahan pada beberapa kota di Indonesia.
6. Menampilkan hasil analisis menggunakan perangkat lunak.

1.4 Batasan Masalah

Batasan-batasan masalah untuk penelitian ini adalah sebagai berikut:

1. Data terkait wilayah desa, kelurahan, kota, dan provinsi di Indonesia yang diperoleh dari situs Badan Informasi Geospasial Indonesia⁶.
2. Penelitian dilakukan pada sebuah kluster yang memiliki jumlah 11 komputer yang terdiri dari 1 buah komputer *master* dan 10 buah komputer *slave* dan masing-masing komputer memiliki memori minimum 8GB dan *core processor* berjumlah 6 buah.
3. Perhitungan luas RTH berdasarkan data citra satelit yang disediakan oleh situs ArcGIS⁷.
4. Perhitungan luas RTH dilakukan pada wilayah kelurahan di beberapa kota/kabupaten di Indonesia menggunakan teknik *pixel based*.

⁶Situs penyedia data geospasial di Indonesia <https://geoservices.big.go.id/>

⁷Situs penyedia data citra satelit ArcGIS <https://www.arcgis.com/>

1.5 Metodologi

Metodologi yang digunakan dalam penyusunan penelitian ini adalah:

1. Mendefinisikan masalah yang ingin diselesaikan
2. Melakukan studi literatur mengenai penelitian sebelumnya.
3. Melakukan studi literatur mengenai RTH.
4. Melakukan studi literatur dan studi eksplorasi mengenai citra satelit.
5. Melakukan studi literatur dan studi eksplorasi mengenai framework Apache Hadoop dan Apache Spark.
6. Mempelajari bahasa pemrograman Scala dan Java beserta *library-library* terkait.
7. Mengumpulkan data citra satelit dan disimpan ke HDFS.
8. Melakukan segmentasi pada citra satelit menggunakan algoritma K-Means.
9. Menghitung luas RTH.
10. Membuat rancangan perangkat lunak.
11. Membangun perangkat lunak.
12. Melakukan pengujian dan eksperimen terhadap perangkat lunak untuk persentase luas area hijau berdasarkan hasil perhitungan menggunakan teknologi *big data*.
13. Menulis dokumen skripsi

1.6 Sistematika Pembahasan

Laporan penelitian tersusun ke dalam enam bab secara sistematis sebagai berikut:

- **Bab 1: Pendahuluan**
Berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika pembahasan.
- **Bab 2: Dasar Teori**
Berisi dasar teori tentang ruang terbuka hijau, ruang terbuka hijau wilayah kelurahan, citra satelit, teknologi *big data*, algoritma klasterisasi, algoritma K-Means, metode evaluasi elbow, algoritma K-Means paralel, *data lake* dan Base64.
- **Bab 3: Pengumpulan, Penyiapan, dan Eksplorasi data**
Berisi pengumpulan, penyiapan, penyiapan data, pemodelan tahap awal berupa pembuatan fitur citra satelit, klasterisasi menggunakan algoritma K-Means, mengevaluasi model klasterisasi menggunakan metode elbow
- **Bab 4: Analisis Hasil Pengumpulan dan Segmentasi Citra Satelit**
Berisi analisa waktu mengumpulkan data citra satelit, analisa hasil pengumpulan citra, perbedaan kontras warna citra satelit, analisa waktu memproses data citra satelit, membandingkan hasil antara yang menggunakan teknik terdistribusi dan tidak terdistribusi, evaluasi luas wilayah keluarahan dengan luas sebenarnya.
- **Bab 5: Pembangunan Perangkat Lunak**
Berisi perancangan *data lake*, struktur pada *data lake*, arsitektur perangkat lunak, hubungan perangkat lunak dengan *data lake*, perancangan fitur dan kebutuhan perangkat lunak, perancangan antarmuka, diagram kelas perangkat lunak, dan hasil implementasi perangkat lunak.
- **Bab 6: Kesimpulan dan Saran**
Berisi kesimpulan dan saran dari penelitian yang telah dilakukan.

