

**CLUSTERING PEMANFAATAN LAHAN BERDASARKAN LUASAN
DAN RENCANA TATA RUANG WILAYAH MENGGUNAKAN
METODE K-MEANS**

SKRIPSI

**Oleh :
SATRIA RAMADHAN DIMASTORY
NIM. 19650083**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

**CLUSTERING PEMANFAATAN LAHAN BERDASARKAN LUASAN
DAN RENCANA TATA RUANG WILAYAH MENGGUNAKAN
METODE K-MEANS**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
SATRIA RAMADHAN DIMASTORY
NIM. 19650083

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

HALAMAN PERSETUJUAN

CLUSTERING PEMANFAATAN LAHAN BERDASARKAN LUASAN DAN RENCANA TATA RUANG WILAYAH MENGGUNAKAN METODE K-MEANS

SKRIPSI

Oleh :
SATRIA RAMADHAN DIMASTORY
NIM. 19650083

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal : 25 Juni 2026

Pembimbing I,



Dr. Agung Teguh Wibowo A, M.T
NIP. 19860301 202321 1 016

Pembimbing II,



Roro Inda Melani, M.T, M.Sc
NIP. 19780925 200501 2 008

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyanto, M. Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

CLUSTERING PEMANFAATAN LAHAN BERDASARKAN LUASAN DAN RENCANA TATA RUANG WILAYAH MENGGUNAKAN METODE K-MEANS

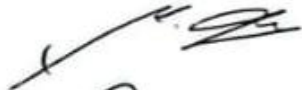
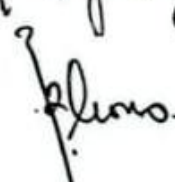
SKRIPSI

Oleh :
SATRIA RAMADHAN DIMASTORY
NIM. 19650083

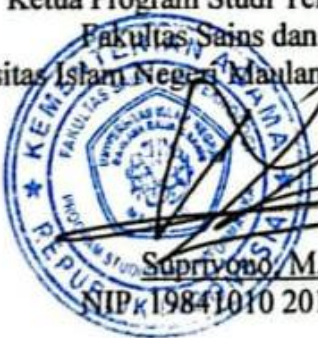
Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal : 25 Juni 2026

Susunan Dewan Penguji

Ketua Penguji	: <u>Dr. M. Ainul Yaqin, M.Kom</u> NIP. 19761013 200604 1 004
Anggota Penguji I	: <u>A'la Syaui, M.Kom</u> NIP. 19771201 200801 1 007
Anggota Penguji II	: <u>Dr. Agung Teguh Wibowo A, M.T</u> NIP. 19860301 202321 1 016
Anggota Penguji III	: <u>Roro Inda Melani, M.T, M.Sc</u> NIP. 19780925 200501 2 008

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono M. Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Satria Ramadhan Dimastory
NIM : 19650083
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Clustering Pemanfaatan Lahan Berdasarkan Luasan
Dan Rencana Tata Ruang Wilayah Menggunakan
K-Means

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 25 Juni 2026
Yang membuat pernyataan,



Satria Ramadhan Dimastory
NIM.19650083

MOTTO

...Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia lain...

HALAMAN PERSEMBAHAN

Syukur mendalam penulis panjatkan ke hadirat Allah SWT atas segala rahmat, hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini.

Dengan rasa hormat dan cinta, karya ini penulis persembahkan kepada:

Kedua orang tua tercinta, Bapak Drs. R. Agoes Trijono dan Ibu Minarti, atas cinta, doa, dan pengorbanan yang tiada henti dalam mengiringi setiap langkah kehidupan penulis.

Dosen pembimbing dan seluruh sivitas akademika, atas ilmu, bimbingan, dan dukungan yang begitu berarti selama proses pendidikan penulis.

Teman-teman seperjuangan, yang telah menjadi bagian penting dalam perjalanan ini melalui kebersamaan, dukungan, dan semangat juang bersama.

Dan untuk diri sendiri, terima kasih telah bertahan, terus berusaha.

Semoga karya ini menjadi langkah awal yang bermanfaat bagi banyak pihak.

KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh

Segala puji hanya bagi Allah SWT yang telah melimpahkan rahmat, hidayah, dan kekuatan-Nya, sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Clustering Pemanfaatan Lahan Berdasarkan Luasan dan Rencana Tata Ruang Wilayah Menggunakan Metode K-Means" dengan baik dan lancar. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, yang telah membawa umat manusia dari zaman kegelapan menuju cahaya Islam yang penuh petunjuk.

Skripsi ini merupakan hasil dari perjalanan panjang dalam menyelesaikan tugas akhir perkuliahan sebagai salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Dalam proses penyusunan skripsi ini, penulis menerima banyak bantuan, dukungan, serta bimbingan dari berbagai pihak. Oleh karena itu, penulis menyampaikan rasa terima kasih yang tulus kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M. Si., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M. Kes., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M. Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Agung Teguh Wibowo Almais, M.T., selaku Dosen Pembimbing I yang telah dengan sabar dan tulus membimbing penulis sepanjang perjalanan skripsi

ini. Bimbingan, arahan, dan dukungan yang diberikan tidak hanya sekadar ilmiah, tetapi juga penuh dengan inspirasi dan semangat yang membangkitkan motivasi penulis untuk terus berusaha.

5. Roro Inda Melani, M.T., M.Sc., selaku Dosen Pembimbing II yang telah membimbing dan memberikan arahan kepada penulis dalam proses penulisan skripsi, termasuk dalam integrasi nilai-nilai keislaman ke dalam karya ilmiah ini.
6. Dr. M. Ainul Yaqin, M.Kom., selaku Dosen Penguji I, dan A'la Syauqi, M.Kom., selaku Dosen Penguji II, yang telah menguji serta memberikan saran dan masukan yang sangat berharga, sehingga penulis dapat menyempurnakan skripsi ini dengan lebih baik.
7. Segenap dosen, tenaga administrasi dan seluruh jajaran staf di Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang, yang telah memberikan ilmu, pelayanan, dan dukungan selama penulis menempuh studi.
8. Kedua orang tua tercinta, Bapak Drs. R. Agoes Trijono dan Ibu Minarti, yang telah memberikan cinta, doa, serta dukungan yang tiada henti. Tanpa kehadiran, kasih sayang, dan pengorbanan mereka, penulis tidak akan mampu mencapai tahap ini. Mereka adalah sumber kekuatan, inspirasi, dan semangat yang tak pernah padam di setiap langkah perjuangan penulis. Terima kasih atas segalanya, atas setiap doa yang dipanjatkan, dan atas kasih tanpa syarat yang selalu menyertai.

9. Istri dan putri penulis, yang selalu menjadi sumber kekuatan dalam hidup penulis. Terima kasih atas kesabaran, doa, serta dukungan yang tak pernah surut selama penulis menjalani proses penyusunan skripsi ini. Kehadiran kalian adalah anugerah terindah yang senantiasa memberi semangat di setiap langkah perjuangan ini.

Malang, 25 Juni 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	1
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
ABSTRAK	xv
ABSTRACT	xvi
الملخص	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
BAB II STUDI PUSTAKA.....	7
2.1 Penelitian Terkait	7
2.2 Konsep Dasar Pemanfaatan Lahan.....	14
2.3 K-Means Clustering	17
BAB III DESAIN DAN IMPLEMENTASI.....	19
3.1 Data dan Sumber Data	19
3.1.1 Jenis Data	19
3.1.2 Sumber Data.....	20
3.2 Alat dan Bahan	21
3.2.1 Perangkat Keras	21
3.2.2 Perangkat Lunak	22
3.3 Pengumpulan Data.....	23
3.3.1 Data Preparation.....	24
3.3.2 Transformasi Data.....	25
3.4 Desain Penelitian	26
3.5 Clustering (K-Means)	28
3.5.1 Konsep K-Means	28
3.5.2 Tahapan Clustering	29
3.5.3 Implementasi Model K-Means Standar	30
3.5.4 Implementasi Model K-Means Modifikasi	31
3.5.5 Analisis Hasil Cluster	32
BAB IV UJI COBA DAN PEMBAHASAN	34
4.1 Deskripsi Data.....	34

4.2 <i>Preprocessing</i> Data	35
4.3 <i>Clustering K-Means</i>	38
4.3.1 Pemilihan <i>K</i> dan Hasil <i>Clustering</i> pada Model <i>K-Means</i> Standar.....	38
4.3.2 Pemilihan <i>K</i> dan Hasil <i>Clustering</i> pada Model <i>K-Means</i> Modifikasi	43
4.3.3 Perbandingan Model <i>K-Means</i>	47
4.4 Intepretasi Hasil dengan Topik	49
BAB V KESIMPULAN DAN SARAN.....	54
5.1 Kesimpulan	54
5.2 Saran	57
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3.1 Desain Penelitian	25
Gambar 4.1 Hasil Clustering Model K-Means Standar (K=2)	35
Gambar 4.2 Hasil Clustering Model K-Means Standar (K=3)	36
Gambar 4.3 Hasil Clustering Model K-Means Standar (K=4)	37
Gambar 4.4 Hasil Clustering Model K-Means Standar (K=5)	38
Gambar 4.5 Hasil Clustering Model K-Means Modifikasi (K=2)	40
Gambar 4.6 Hasil Clustering Model K-Means Modifikasi (K=3)	41
Gambar 4.7 Hasil Clustering Model K-Means Modifikasi (K=4)	42
Gambar 4.8 Hasil Clustering Model K-Means Modifikasi (K=5)	43

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	11
Tabel 3.1 Data Collection	19
Tabel 3.2 Mapping Label Encoding Kolom RTRW	24
Tabel 4.1 Statistik Deskriptif Dataset	31
Tabel 4.2 Contoh Data Setelah Proses Encoding	33
Tabel 4.3 Hasil Clustering Model K-Means Standar (K=2)	35
Tabel 4.4 Hasil Clustering Model K-Means Standar (K=3)	36
Tabel 4.5 Hasil Clustering Model K-Means Standar (K=4)	39
Tabel 4.8 Hasil Clustering Model K-Means Modifikasi (K=2)	40
Tabel 4.9 Hasil Clustering Model K-Means Modifikasi (K=3)	41
Tabel 4.10 Hasil Clustering Model K-Means Modifikasi (K=4)	42
Tabel 4.11 Hasil Clustering Model K-Means Modifikasi (K=5)	43
Tabel 4.12 Evaluasi Silhouette Score Model K-Means Modifikasi	45

ABSTRAK

Dimastory, Satria. 2026. **Clustering Pemanfaatan Lahan Berdasarkan Luasan Dan Rencana Tata Ruang Wilayah Menggunakan Metode K-Means**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Agung Teguh Wibowo A, M.T (II) Roro Inda Melani, M.T, M.Sc.

Kata kunci: K-Means, clustering, pemanfaatan lahan, RTRW, Desa Sambiresik.

Pengelolaan lahan merupakan aspek penting dalam pembangunan wilayah yang berkelanjutan, terutama di tengah keterbatasan lahan dan meningkatnya kebutuhan sektoral. Penelitian ini bertujuan untuk mengelompokkan pemanfaatan lahan berdasarkan luasan dan Rencana Tata Ruang Wilayah (RTRW) menggunakan metode K-Means Clustering. Lokasi penelitian difokuskan di Desa Sambiresik, Kabupaten Kediri, dengan data yang diperoleh dari Badan Pertanahan Nasional (ATR/BPN) setempat, meliputi Nomor Identifikasi Bidang Tanah (NIB), luas tanah, dan informasi RTRW. Hasil analisis menunjukkan bahwa metode K-Means efektif dalam mengidentifikasi pola pemanfaatan lahan yang lebih terarah dan efisien. Clustering ini diharapkan dapat menjadi dasar pertimbangan bagi instansi seperti BPN dan Dinas Perumahan dan Kawasan Permukiman (Perkim) dalam mempermudah proses perizinan, seperti PBG (Pendirian Bangunan Gedung). Penelitian ini menegaskan pentingnya pendekatan digital dan berbasis data dalam mewujudkan pengelolaan lahan yang berkelanjutan, terintegrasi, dan mendukung kebijakan pembangunan nasional..

ABSTRACT

Dimastory, Satria. 2026. **Clustering Pemanfaatan Lahan Berdasarkan Luasan Dan Rencana Tata Ruang Wilayah Menggunakan Metode K-Means**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Promotor: (I) Dr. Agung Teguh Wibowo A, M.T (II) Roro Inda Melani, M.T, M.Sc.

Key words: K-Means, clustering, land use, spatial planning, Sambiresik Village.

Land management is a crucial aspect of sustainable regional development, especially amidst limited land availability and increasing sectoral demands. This study aims to cluster land use based on land area and Spatial Planning (RTRW) using the K-Means Clustering method. The research was conducted in Sambiresik Village, Kediri Regency, utilizing data obtained from the local National Land Agency (ATR/BPN), including Land Parcel Identification Numbers (NIB), land area, and spatial planning information. The analysis results indicate that the K-Means method is effective in identifying more targeted and efficient land use patterns. This clustering approach is expected to serve as a reference for institutions such as BPN and the Housing and Settlement Area Agency (Perkim) in facilitating the building permit (PBG) process. The study emphasizes the importance of digital and data-driven approaches in achieving sustainable, integrated land management that supports national development policies.

الملخص

ديماستوري، ساتريا. ٢٠٢٦. تجميع استخدامات الأراضي بناءً على المساحة وخطة استعمالات وتخطيط المناطق باستخدام طريقة رسالة تخرج. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية. **K-Means** الحكومية مالانج. المشرفان: (١) الدكتور أغونغ تيغوه ويوو أ، ماجستير في الهندسة، (٢) رورو إندا ميلاني، ماجستير في الهندسة، وماجستير في العلوم. التجميع، استخدام الأراضي، التخطيط المكاني، قرية سامبيريسيك، **K-Means**: الكلمات المفتاحية

إدارة الأراضي تُعد جانبًا حيويًا في تنمية المناطق بشكل مستدام، خاصةً في ظل محدودية الأراضي وازدياد الطلب القطاعي. تهدف باستخدام (RTRW) هذه الدراسة إلى تصنيف استخدام الأراضي بناءً على المساحة وخطة الاستخدامات المكانية أُجريت الدراسة في قرية سامبيريسيك، محافظة كيديري، بالاعتماد على بيانات تم الحصول. طريقة التجميع **K-Means**، ومساحة الأرض، (NIB) والتي تشمل رقم تعريف قطعة الأرض، (ATR/BPN) عليها من الوكالة الوطنية للأراضي فعالة في تحديد أنماط استخدام **K-Means** ومعلومات خطة الاستخدامات المكانية. أظهرت نتائج التحليل أن طريقة الأراضي بشكل أكثر دقة وكفاءة. من المتوقع أن يُستخدم هذا التصنيف كمرجع للجهات المعنية مثل وكالة الأراضي تؤكد هذه (PBG) لتسهيل عملية الحصول على تصاريح البناء (Perkim) الوطنية وهيئة الإسكان والمناطق السكنية الدراسة على أهمية اعتماد النهج الرقمي والمبني على البيانات لتحقيق إدارة مستدامة ومتكاملة للأراضي تدعم السياسات الوطنية للتنمية

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengelolaan lahan merupakan aspek krusial dalam pembangunan suatu wilayah yang berkelanjutan. Dalam konteks pembangunan nasional, ketersediaan lahan yang terbatas dan meningkatnya harga tanah telah menjadi tantangan serius bagi pemerintah dalam melaksanakan pembangunan sektoral. keterbatasan lahan menyulitkan pelaksanaan pembangunan[1], sementara menekankan bahwa pengelolaan lahan yang efektif adalah kunci untuk memastikan pemanfaatan sumber daya alam yang berkelanjutan[2].

Pemanfaatan lahan tidak hanya memengaruhi keberlanjutan lingkungan, tetapi juga memiliki dampak yang signifikan terhadap pertumbuhan ekonomi suatu wilayah[3]. Untuk mengatasi tantangan ini, pemerintah telah mengeluarkan berbagai kebijakan, seperti Perpres no.18 tahun 2020, yang menuntut efisiensi pengelolaan lahan, terutama dalam sektor produksi pangan dan pemukiman, serta peningkatan pelayanan pertanahan melalui pendekatan digital.

Meskipun demikian, pengelolaan lahan yang efisien tidaklah mudah. Kompleksitas dalam pengelolaan tanah sering menjadi hambatan bagi pemilik lahan dalam mengoptimalkan penggunaan lahan mereka. Dalam konteks ini, clustering pemanfaatan lahan menjadi solusi yang relevan dan penting. Metode clustering memungkinkan pemilik lahan untuk mengelompokkan lahan mereka

berdasarkan karakteristik tertentu, sehingga memungkinkan pengelolaan yang lebih terarah dan efektif.

Dengan demikian, penelitian ini bertujuan untuk mengaplikasikan metode clustering, khususnya metode K-Means, dalam mengelompokkan pemanfaatan lahan berdasarkan luasannya. Selain itu, penelitian ini juga akan mempertimbangkan penetapan Rencana Tata Ruang dan Wilayah (RTRW) sebagai faktor penentu dalam rekomendasi pemanfaatan lahan berdasarkan cluster-nya. Dengan adanya penelitian ini, diharapkan pengelolaan lahan akan menjadi lebih terarah, efisien, dan berkelanjutan, sehingga dapat mendukung pembangunan nasional yang inklusif dan berkelanjutan.

Segala kebijakan yang telah diatur dan disusun oleh pemerintah hendaknya telah menjadi hal yang baik untuk dilaksanakan dan ditaati. Penetapan aturan agraria pasti dimaksudkan untuk memenuhi kebutuhan masyarakat, meningkatkan kesejahteraan ekonomi, serta kelestarian lingkungan. Sebagai warga negara yang baik hendaknya kita menaati segala aturan maupun kebijakan yang telah ditetapkan. Hal ini sejalan dengan Firman Allah SWT dalam Q.S Al-A'la ayat 7-9:

إِنَّهُ يَعْلَمُ الْجَهْرَ وَمَا يَخْفَىٰ وَنُبَشِّرُكَ لِيُسْرَىٰ فَذَكِّرْ إِنَّ نَعْمَ الذِّكْرَىٰ

“Sesungguhnya Dia mengetahui yang terang dan yang tersembunyi. Kami akan melapangkan bagimu jalan kemudahan (dalam segala urusan). Maka, sampaikanlah peringatan jika peringatan itu bermanfaat” (Q.S Al ‘Alaq: 7-9).

Berdasarkan tinjauan tafsir, seperti dalam Tafsir Ibnu Katsir, ayat ini tidak hanya menegaskan kemahatahuan Allah atas segala tindakan lahiriah maupun batiniah manusia, tetapi juga memberikan jaminan bahwa Allah akan memberikan

jalan kemudahan (al-yusra) bagi hamba-Nya yang berupaya melakukan kebaikan. Lebih lanjut, perintah “maka sampaikanlah peringatan” (fadzakir) mengandung makna bahwa setiap ilmu, informasi, maupun aturan harus disampaikan demi mendatangkan maslahat (kebaikan) bagi umat.

Dalam konteks penetapan aturan agraria dan pengelolaan lahan, penyampaian informasi yang akurat, seperti pemetaan dan hasil clustering tata ruang dalam penelitian ini merupakan bentuk pengejawantahan dari upaya saling memberi manfaat. Dengan menyajikan data yang objektif sebagai bahan "peringatan" atau evaluasi bagi pemangku kebijakan dan masyarakat, diharapkan tercipta kepatuhan terhadap regulasi yang ada. Dengan saling memperingatkan dan mengedepankan kemaslahatan bersama, semoga kita selalu diberikan kemudahan oleh Allah SWT dalam melaksanakan pembangunan yang tertib dan berkelanjutan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana K-Means Clustering dapat memenuhi kebutuhan pengelompokan data pemanfaatan lahan yang besar, beragam, dan kompleks di Desa Sambiresik, sehingga pola pemanfaatan lahan berdasarkan luasan bidang tanah dan zonasi RTRW dapat dipahami dan diklasifikasikan secara objektif?

2. Bagaimana hasil klasterisasi pemanfaatan lahan yang terbentuk dapat dimanfaatkan sebagai acuan berbasis data bagi Badan Pertanahan Nasional (BPN) dan Dinas Perumahan dan Kawasan Pemukiman (Perkim) dalam mendukung evaluasi kesesuaian pemanfaatan lahan terhadap RTRW, baik dalam mempercepat proses perizinan PBG maupun dalam mengantisipasi tekanan alih fungsi lahan yang semakin masif akibat urbanisasi dan pertumbuhan penduduk?

1.3 Batasan Masalah

Berikut adalah batasan masalah yang diterapkan dalam penelitian ini:

1. Lokasi Penelitian

Lokasi penelitian ini berlokasi di Desa Sambiresik, yang terletak di Kecamatan Sambiresik, Kabupaten Kediri. Penentuan Desa Sambiresik sebagai lokasi penelitian didasarkan pada ketersediaan data administrasi yang secara rutin diperbarui dan terstruktur dengan baik di Badan Pertanahan Nasional (ATR/BPN) Kabupaten Kediri. Keberadaan data yang terkini dan terstruktur tersebut dianggap mampu mendukung kelancaran serta ketepatan pelaksanaan penelitian. Desa Sambiresik dipandang sebagai pilihan yang optimal untuk memastikan keakuratan dan kredibilitas data yang digunakan dalam proses analisis.

2. Data Penelitian

Data penelitian ini diperoleh dari Badan Pertanahan Nasional Kabupaten Kediri yang meliputi Nomor Identifikasi Bidang Tanah (NIB), luas tanah dan Rencana Tata Ruang Wilayah (RTRW). Keputusan untuk menggunakan data dari

lembaga tersebut didasarkan pada kehandalan dan kredibilitasnya dalam menyajikan informasi yang akurat dan terpercaya terkait wilayah yang diteliti.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini yakni mengetahui efektivitas metode k-means dalam clustering pemanfaatan lahan berdasarkan luasan dan rencana tata ruang wilayah yang dapat menjadi bahan pertimbangan dalam melakukan rencana pemanfaatan lahan.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat yang signifikan bagi berbagai pihak terkait, termasuk pemilik lahan, pengembang, dan dinas terkait, dalam konteks pengelolaan lahan yang efektif dan berkelanjutan. Berikut adalah beberapa manfaat yang diharapkan dari penelitian ini:

1. Informasi yang Tepat untuk Pengambilan Keputusan: Hasil dari penelitian ini akan memberikan informasi yang tepat mengenai potensi pemanfaatan lahan berdasarkan hasil clustering. Informasi ini akan menjadi dasar yang kuat bagi pemilik lahan dalam mengambil keputusan terkait penggunaan lahan mereka, serta bagi pengembang dalam merencanakan pembangunan yang lebih terarah.
2. Optimalisasi Penggunaan Lahan: Dengan adanya hasil clustering, pemilik lahan dapat mengoptimalkan penggunaan lahan mereka sesuai dengan regulasi tata ruang yang berlaku. Hal ini akan memungkinkan penggunaan

lahan yang lebih efisien dan sesuai dengan kebutuhan produksi pangan dan pemukiman masyarakat.

3. Peningkatan Kualitas Lingkungan: Pengelolaan lahan yang lebih terarah dan efektif akan berdampak positif pada kualitas lingkungan. Dengan adanya informasi yang akurat mengenai potensi pemanfaatan lahan, penelitian ini dapat membantu dalam menjaga keberlanjutan lingkungan dan mengurangi dampak negatif terhadap ekosistem alam.
4. Kontribusi pada Pembangunan Berkelanjutan: Dengan memberikan informasi yang relevan dan akurat mengenai pemanfaatan lahan, penelitian ini diharapkan dapat memberikan kontribusi pada pembangunan berkelanjutan di tingkat lokal maupun nasional. Dengan demikian, penelitian ini tidak hanya akan memberikan manfaat bagi pemilik lahan secara individu, tetapi juga bagi masyarakat luas dan pembangunan negara secara keseluruhan.

Dengan manfaat yang dihasilkan dari penelitian ini, diharapkan akan tercipta pengelolaan lahan yang lebih efisien, berkelanjutan, dan berdampak positif bagi pembangunan nasional.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian Pratama Siregar P (2022) membahas implementasi algoritma K-Means untuk pengelompokan distribusi sosial ekonomi masyarakat berdasarkan demografi kependudukan. Penelitian ini bertujuan untuk mengidentifikasi pola distribusi sosial ekonomi di berbagai wilayah menggunakan data demografi. Algoritma K-Means digunakan untuk mengelompokkan data sehingga memungkinkan pemerintah dan pembuat kebijakan untuk memahami dan mengatasi ketidaksetaraan sosial ekonomi dengan lebih efektif. Hasil penelitian menunjukkan bahwa pengelompokan ini dapat membantu dalam merumuskan kebijakan yang lebih tepat sasaran untuk pembangunan sosial ekonomi.

Penelitian Syaputri D (2021) mengaplikasikan algoritma K-Means untuk mengelompokkan data demografi kependudukan guna mengidentifikasi distribusi sosial ekonomi masyarakat. Dengan menggunakan data demografi dari berbagai wilayah, algoritma K-Means berhasil mengelompokkan masyarakat ke dalam beberapa kelompok berdasarkan karakteristik sosial ekonomi mereka. Hasil dari penelitian ini diharapkan dapat digunakan untuk merancang kebijakan yang lebih efektif dalam meningkatkan kesejahteraan sosial ekonomi masyarakat.

Penelitian Ilham A (2023) mengusulkan model kombinasi Random Undersampling berbasis K-Means Cluster (RUS-KMC) dan pendekatan meta-learning (RFM) untuk klasifikasi status pembangunan desa di Indonesia. Model ini digunakan untuk mengatasi masalah kelas tidak seimbang dalam dataset Indeks

Desa Membangun (IDM). Hasil evaluasi menunjukkan bahwa kombinasi RUS-KMC dan RFM mampu meningkatkan kinerja klasifikasi dibandingkan dengan metode sebelumnya, dengan nilai AUC, ACC, PR, RC, dan Gm masing-masing mencapai 95.50%, 95.52%, 95.5%, 95.5%, dan 92.95%. Model ini efektif dalam meningkatkan akurasi dan penanganan kelas tidak seimbang dalam klasifikasi status pembangunan desa.

Penelitian Almais A (2023) menggunakan teknik clustering data untuk menentukan label dari hasil nilai numerik dari teknik PCA yang menggunakan 216 data kerusakan pada sektor pasca-bencana alam yang telah melalui proses analisis dan memiliki nilai pada setiap kriteria. Kemudian PCA memproses 216 data tersebut untuk menghasilkan nilai numerik berdasarkan hasil grafik PCA, hasil nilai numerik ini dikelompokkan untuk menentukan jenis label untuk tingkat kerusakan pada sektor setelah bencana alam. Dalam geofisika, penggunaan PCA berfungsi untuk mengklasifikasikan fasies seismik dengan memberikan label dan ukuran secara otomatis sehingga peta seismik dapat dikelompokkan dan disajikan secara rapi dan jelas.

Penelitian Fithriyah M (2021) mengatur segmentasi produk berdasarkan karakteristiknya, sehingga memudahkan pemilik bisnis dalam mengelola stok dan memenuhi permintaan pelanggan dengan tepat waktu dengan model RFM digunakan untuk menganalisis data transaksi produk guna mendapatkan nilai indeks untuk setiap produk. Seluruh data produk dikelompokkan menggunakan algoritma k-means clustering berdasarkan nilai indeksnya. Proses clustering dilakukan dengan variasi jumlah cluster (2, 3, 4, 5, dan 6) untuk mencari jumlah cluster yang

optimal dalam mengelompokkan data produk. Hasil clustering dievaluasi menggunakan metrik silhoutte untuk menilai kualitasnya. Hasil pengujian menunjukkan nilai silhoutte tertinggi adalah 0,4314 pada clustering kedua dengan jumlah cluster 3.

Tabel 2.1 Penelitian Terkait

No	Penulis dan Tahun Penelitian	Judul	Objek dan Metode	Hasil Penelitian
1.	Siregar P.P (2022)	Penerapan Metode K-Means Dalam Mengelompokkan Persebaran Lahan Kritis Di Indonesia Berdasarkan Provinsi	Menggunakan algoritma K- Means untuk mengelompokkan data demografi kependudukan dan mengidentifikasi pola distribusi sosial ekonomi.	Algoritma K- Means efektif dalam mengelompokkan masyarakat berdasarkan karakteristik sosial ekonomi, yang dapat membantu dalam merumuskan kebijakan sosial ekonomi yang lebih tepat sasaran.

No	Penulis dan Tahun Penelitian	Judul	Objek dan Metode	Hasil Penelitian
2.	Syaputri D (2021)	Implementasi Algoritma K- Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan	Menggunakan data demografi dan algoritma K- Means untuk pengelompokan sosial ekonomi masyarakat.	Hasil pengelompokan membantu dalam perencanaan kebijakan sosial ekonomi yang lebih efektif.
3.	Ilham A (2023)	k-Means Cluster- based Random Undersampling and Meta- Learning Approach for Village Development	Model kombinasi Random Undersampling berbasis K-Means Cluster (RUS- KMC) dan meta- learning (RFM) untuk klasifikasi status	Mengatasi masalah kelas tidak seimbang dalam dataset Indeks Desa Membangun (IDM) menggunakan kombinasi RUS- KMC dan RFM.

No	Penulis dan Tahun Penelitian	Judul	Objek dan Metode	Hasil Penelitian
		Status Classification	pembangunan desa.	
4.	Almais A (2023)	Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post- Natural Disasters	Principal Component Analysis (PCA)	Pembuatan label dari suatu kumpulan data menjadi data yang baik dan mudah untuk digunakan dalam proses pembelajaran mesin atau deep learning menggunakan teknik PCA.
5.	Fithriyah M (2021)	K-Means Clustering Untuk Segmentasi Produk Berdasarkan	Menggunakan algoritma K- Means untuk clustering dan	Ditemukan 3 kelompok produk dengan karakteristik yang berbeda.

No	Penulis dan Tahun Penelitian	Judul	Objek dan Metode	Hasil Penelitian
		Analisis Recency, Recency, Recency (RFM) pada Data Transaksi Penjualan	RFM untuk klasifikasi	Kelompok C1 terdiri dari produk dengan kinerja recency, frequency, dan monetary yang baik, C2 terdiri dari produk dengan kinerja yang buruk, dan C3 terdiri dari produk dengan kinerja yang cukup baik.
6.	Troccoli (2023)	K-means clustering using principal component analysis to automate label	K-Means & PCA	Menginterpretasi dua permukaan stratigrafi penting, yaitu Permukaan Transgresi Maksimum (MTS)

No	Penulis dan Tahun Penelitian	Judul	Objek dan Metode	Hasil Penelitian
		organization in multi-attribute seismic facies analysis		dan Permukaan Regresi Maksimum (MRS), dan mengekstrak atribut-atribut relevan untuk eksperimen pengelompokan.
7.	Utnasari I (2021)	Analisis Clustering Dengan K-Means Untuk Pengelompokkan Penjualan Produk Pada Hotel Newton	Menggunakan algoritma K- means clustering dalam data mining	Hasil clustering 3 kelompok barang yang sangat laku terjual, barang yang cukup laku terjual, dan barang yang tidak laku terjual.
Penelitian yang akan dilakukan				

No	Penulis dan Tahun Penelitian	Judul	Objek dan Metode	Hasil Penelitian
1.	Satria Ramadhan Dimastory	Clustering Pemanfaatan Lahan Berdasarkan Luasan Dan Rencana Tata Ruang Wilayah Menggunakan Metode K-Means	Menggunakan algoritma K- Means Clustering.	Hasil clustering yang dapat dimanfaatkan sebagai proyeksi pemanfaatan lahan.

2.2 Konsep Dasar Pemanfaaaatan Lahan

Pemanfaatan lahan merupakan kebijakan agraria yang mengatur penggunaan lahan. Berikut adalah beberapa konsep dasar pemanfaatan lahan dengan referensi Permen ATR/BPN no.13 tahun 2021:

1. Penggunaan Lahan Pertanian: Lahan yang digunakan untuk produksi pangan, tanaman, dan perkebunan. Pengelolaan yang baik terhadap lahan pertanian sangat penting untuk memastikan ketahanan pangan dan kesejahteraan petani.

2. **Penggunaan Lahan Non-Pertanian:** Lahan yang digunakan untuk pemukiman, industri, dan komersial. Pengaturan yang bijaksana dalam penggunaan lahan non-pertanian penting untuk menjaga keseimbangan antara pembangunan ekonomi dan pelestarian lingkungan.
3. **Regulasi Tata Ruang Wilayah (RTRW):** RTRW berfungsi mengatur serta menentukan terkait zonasi dan perizinan penggunaan lahan berdasarkan karakteristik dan potensi wilayah. Regulasi tersebut membantu mengoptimalkan pemanfaatan lahan dan mengurangi konflik kepentingan.
4. **Lahan Sawah Dilindungi (LSD):** LSD merupakan area yang dilindungi untuk menjaga keberlanjutan produksi pangan. Perlindungan ini meliputi pemeliharaan kesuburan tanah, pengendalian konversi lahan, dan pemulihan lahan yang rusak.
5. **Prinsip Keberlanjutan:** Pemanfaatan lahan yang berkelanjutan memperhitungkan aspek lingkungan, ekonomi, dan sosial. Prinsip-prinsip seperti konservasi tanah, biodiversitas, dan pembangunan ekonomi lokal telah menjadi fokus dalam kebijakan agraria.
6. **Partisipasi Masyarakat:** Masyarakat harus terlibat dalam proses pengembangan dan pengambilan keputusan memastikan kebijakan yang dihasilkan sesuai dengan kebutuhan lokal dan negara.
7. **Pengelolaan Risiko:** Pengelolaan risiko melibatkan mitigasi dan adaptasi terhadap ancaman seperti bencana alam dan perubahan iklim.

8. Pengawasan dan Penegakan Hukum: Mekanisme pengawasan dan penegakan hukum yang kuat diperlukan untuk memastikan kepatuhan terhadap regulasi dan mencegah pelanggaran.

Dengan memperhatikan konsep-konsep ini, kebijakan agraria dapat dirancang untuk mendukung pertumbuhan ekonomi yang berkelanjutan sambil melindungi lingkungan dan meningkatkan kesejahteraan masyarakat, tanpa mengorbankan salah satu aspeknya.

Penelitian ini tidak hanya menekankan pentingnya pendekatan data melalui analisis klaster, tetapi juga mempertimbangkan nilai-nilai moral dan spiritual. Dalam konteks ini, terdapat relevansi kuat antara pengelolaan lahan dan ajaran islam yang menekankan keseimbangan, tanggung jawab, dan larangan melakukan kerusakan. Sebagaimana firman Allah SWT dalam Q.S. Al-Baqarah ayat 60:

وَكُلُوا وَاشْرَبُوا مِنْ رِزْقِ اللَّهِ وَلَا تَعْلَوْا فِي الْأَرْضِ مُفْسِدِينَ

"Dan makan dan minumlah dari rezeki (yang diberikan) Allah, dan janganlah kamu berkeliaran di muka bumi dengan berbuat kerusakan." (Q.S Al Baqarah: 60).

Ayat ini menegaskan bahwa segala bentuk sumber daya alam, termasuk lahan, merupakan rezeki dari Allah SWT yang harus dikelola dengan bijak tanpa menimbulkan kerusakan. Merujuk pada penjelasan para ahli tafsir, seperti dalam Tafsir Al-Misbah, kata *la ta'tsau* berakar dari kata yang bermakna kerusakan yang sangat parah atau eksploitasi yang melampaui batas. Larangan berbuat fasad (kerusakan) di sini bermakna luas, tidak hanya kerusakan fisik lingkungan, tetapi juga perusakan tatanan sosial dan ekosistem akibat keserakahan manusia.

Dalam perspektif penelitian ini, pelanggaran zonasi lahan, alih fungsi Lahan Sawah Dilindungi (LSD) secara serampangan, atau pengelolaan tata ruang yang tidak terkendali dapat dikategorikan sebagai bentuk kerusakan di muka bumi. Oleh karena itu, upaya memetakan dan mengelola lahan sesuai dengan peruntukannya (RTRW) merupakan bentuk pengejawantahan dari rasa syukur atas rezeki Allah, sekaligus ikhtiar untuk menjaga keseimbangan ekologis dan mewujudkan pembangunan yang selaras dengan nilai-nilai syariat Islam.

2.3 K-Means Clustering

Algoritma K-Means Clustering merupakan algoritma yang efisien dan cepat dalam mengelompokkan data. Prinsip kerjanya dimulai dengan pemilihan K centroid, K adalah jumlah cluster yang diinginkan. Setelah itu, nilai centroid diperbarui secara iteratif dengan menghitung jarak antara setiap data dan centroid menggunakan metrik Euclidean (Siregar P.P, 2022). Data kemudian diklasifikasikan berdasarkan kedekatannya dengan centroid yang sesuai. Proses ini diulang sampai nilai centroid stabil, yang menandakan konvergensi algoritma.

K-Means Clustering adalah teknik pengelompokan data non-hirarkis yang mempartisi data menjadi dua atau lebih kelompok sehingga data dengan karakteristik serupa ditempatkan dalam satu kelompok yang sama. Langkah-langkah umum dalam melakukan clustering menggunakan algoritma K-Means adalah sebagai berikut:

Tentukan nilai k sebagai jumlah cluster yang ingin dibentuk.

Membangkitkan k centroid (titik pusat cluster) awal secara random/acak, kemudian untuk menghitung centroid cluster ke-i berikutnya, digunakan rumus sebagai berikut :

$$v = \sqrt{\frac{\sum_{i=1}^n x_i}{n}}$$

Dimana :

i : 1,2,3,...n

v : centroid pada cluster

x_i : objek ke-i

n : banyaknya jumlah objek yang menjadi anggota cluster

Hitung jarak setiap data ke masing-masing centroid menggunakan rumus.

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan :

dij : Jarak objek antara objek x dan y

n : Jumlah Atribut

x_i : Objek Data

y_i : Data Cluster

Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya (C).

Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.

BAB III

METODE PENELITIAN

Bab ini menjabarkan secara teknis dan metodologis seluruh tahapan yang ditempuh dalam penelitian ini, mulai dari pengumpulan data, pra-pemrosesan, transformasi data, implementasi algoritma K-Means Clustering, hingga evaluasi kualitas hasil clustering. Setiap tahapan dijelaskan secara rinci agar penelitian ini dapat direproduksi dan dipahami oleh pembaca dengan latar belakang ilmu komputer maupun bidang terkait.

3.1 Data dan Sumber Data

Penjelasan pada bagian ini difokuskan pada karakteristik tipe data beserta atribut-atribut teknis yang digunakan sebagai variabel clustering, serta asal-usul pemerolehan data guna memastikan validitas dan akurasi informasi. Pemaparan ini sangat penting untuk memberikan gambaran komprehensif mengenai kualitas dan struktur dataset sebelum melangkah ke tahap komputasi lebih lanjut.

3.1.1 Jenis Data

Data yang digunakan dalam penelitian ini merupakan data administratif pertanahan yang bersifat semi-sensitif, mengingat data tersebut memuat informasi identifikasi bidang tanah yang terasosiasi dengan kepemilikan lahan di wilayah tertentu. Data diperoleh dari Badan Pertanahan Nasional (ATR/BPN) Kabupaten Kediri melalui proses koordinasi resmi dengan pihak instansi terkait, dengan pembatasan detail informasi yang diizinkan untuk dipublikasikan sesuai ketentuan Kepala BPN Kabupaten Kediri.

Data yang digunakan mencakup atribut-atribut berikut: Nomor Identifikasi Bidang Tanah (NIB) sebagai kode unik per bidang tanah, Tahun pendaftaran bidang tanah, Luas bidang tanah dalam satuan meter persegi (m^2), Rencana Tata Ruang Wilayah (RTRW) yang menunjukkan zonasi fungsi lahan, dan status Lahan Sawah Dilindungi (LSD). Untuk keperluan klasterisasi, penelitian ini memfokuskan dua fitur utama yaitu LUAS dan RTRW, karena keduanya paling merepresentasikan karakteristik pemanfaatan lahan yang relevan dengan tujuan penelitian.

Tipe data yang digunakan merupakan kombinasi antara data numerik (LUAS dalam m^2) dan data kategorikal (RTRW). Data kategorikal memerlukan proses encoding terlebih dahulu sebelum dapat diproses oleh algoritma K-Means, mengingat algoritma ini hanya dapat beroperasi pada data bertipe numerik.

3.1.2 Sumber Data

Data penelitian ini diperoleh langsung dari Kantor Badan Pertanahan Nasional (ATR/BPN) Kabupaten Kediri, yang merupakan instansi pemerintah berwenang dalam pengelolaan administrasi pertanahan di Indonesia. Data disajikan dalam format file Microsoft Excel (.xlsx) yang mencakup seluruh bidang tanah terdaftar di wilayah Desa Sambiresik, Kecamatan Gampengrejo, Kabupaten Kediri, Jawa Timur.

Total data yang berhasil diperoleh mencakup 2.818 baris data bidang tanah, dengan rentang NIB dari 1781 hingga 4598. Data mencakup bidang tanah yang terdaftar pada dua kategori RTRW utama, yaitu Kawasan Permukiman Perkotaan (2.230 bidang atau sekitar 79,13% dari total) dan Kawasan Tanaman Pangan (586 bidang atau sekitar 20,79% dari total). Desa Sambiresik dipilih sebagai lokasi

penelitian karena ketersediaan data administrasi yang terkini, terstruktur, dan dapat dipertanggungjawabkan keakuratannya oleh lembaga berwenang. Data diperoleh dari BPN Kabupaten Kediri yang mencakup atribut-atribut sebagai berikut:

1. Nomor Identifikasi Bidang Tanah (NIB)
2. Tahun berkas (Tahun)
3. Luas bidang tanah (m²)
4. Status Rencana Tata Ruang Wilayah (RTRW)
5. Status Lahan Sawah Dilindungi (LSD)

Tabel 3.1 Data Collection

No.	Kategori RTRW	Jumlah Bidang Tanah
1	Kawasan Permukiman Perkotaan	2.230 bidang (79,13%)
2	Kawasan Tanaman Pangan	588 bidang (20,87%)
Total	-	2.818 bidang (100%)

3.2 Alat dan Bahan

Pelaksanaan komputasi dalam penelitian ini membutuhkan dukungan instrumen yang memadai untuk memastikan kelancaran, efisiensi, serta akurasi dalam memproses dataset spasial berskala besar. Pemilihan alat dan bahan ini secara khusus disesuaikan dengan kebutuhan komputasi K-Means Clustering guna menjamin reproduksibilitas dari eksperimen yang dilakukan.

3.1.1 Perangkat Keras

Penelitian ini dilaksanakan menggunakan perangkat keras berupa laptop dengan spesifikasi teknis sebagai berikut: Prosesor AMD Ryzen 7 4800H dengan kecepatan base clock 2,9 GHz dan boost clock hingga 4,2 GHz (8 core, 16 thread),

memori RAM sebesar 16 GB DDR4 dual-channel, serta penyimpanan SSD berkapasitas 1 TB dengan antarmuka NVMe. Spesifikasi ini memungkinkan eksekusi komputasi clustering pada dataset berukuran ribuan baris secara efisien. Karena komputasi utama dilakukan di Google Colab (berbasis cloud), perangkat keras lokal berfungsi sebagai terminal akses dan media penyimpanan berkas.

3.1.2 Perangkat Lunak

Lingkungan komputasi yang digunakan adalah Google Colaboratory (Google Colab), sebuah platform berbasis cloud yang menyediakan akses gratis ke interpreter Python 3.x dan sumber daya komputasi GPU/CPU. Google Colab dipilih karena kemudahan berbagi kode (notebook-based), tidak memerlukan instalasi perangkat lunak lokal, serta mendukung visualisasi data secara langsung.

Adapun pustaka-pustaka Python yang digunakan dalam penelitian ini adalah sebagai berikut:

- (1) Pandas (pd) – digunakan untuk membaca, memanipulasi, dan menganalisis data tabular dalam format DataFrame. Operasi yang digunakan meliputi pembacaan CSV, seleksi kolom, penghapusan baris duplikat dan nilai kosong, serta penggabungan DataFrame.
- (2) Matplotlib (plt) – pustaka visualisasi data fundamental di Python, digunakan untuk membuat grafik scatter plot dan histogram dalam format yang dapat dikustomisasi.

(3) Seaborn (sns) – pustaka visualisasi berbasis Matplotlib dengan antarmuka lebih tinggi, digunakan untuk membuat scatter plot hasil clustering yang lebih estetik dengan pewarnaan otomatis per klaster.

(4) Scikit-learn – pustaka machine learning Python yang menyediakan implementasi algoritma K-Means (`sklearn.cluster.KMeans`), encoder kategorikal (`sklearn.preprocessing.LabelEncoder`), serta metrik evaluasi (`sklearn.metrics.silhouette_score`).

3.3 Pengumpulan Data

Proses pengumpulan data dilakukan melalui koordinasi langsung dengan Kantor BPN Kabupaten Kediri. Data diterima dalam format file Microsoft Excel (.xlsx) yang kemudian dikonversi menjadi format CSV (Comma-Separated Values) untuk memudahkan pembacaan oleh Pandas di Python. Proses konversi dilakukan dengan membuka file Excel menggunakan Microsoft Excel kemudian menyimpannya kembali dalam format .csv (UTF-8). Data diperoleh langsung dari BPN Kabupaten Kediri dalam bentuk file Excel (.xlsx). Data tersebut berisi informasi administratif bidang tanah dan parameter terkait yang relevan dengan penelitian ini, meliputi:

1. Nomor Identifikasi Bidang Tanah (NIB) sebagai identitas unik.
2. Luasan bidang tanah sebagai atribut utama.
3. Status RTRW yang menunjukkan zonasi fungsi lahan.
4. Status LSD yang menunjukkan apakah bidang tanah tersebut termasuk lahan sawah yang dilindungi.

3.1.3 Data Preparation

Tahap data preparation merupakan langkah fundamental sebelum data dapat digunakan dalam proses clustering. Pada penelitian ini, data preparation mencakup beberapa sub-proses berikut:

Pertama, pemeriksaan dan penanganan nilai kosong (Missing Values). Dengan menggunakan fungsi `data.isnull().sum()` dari Pandas, sistem memeriksa setiap kolom dataset untuk mendeteksi keberadaan nilai NaN (Not a Number). Jika ditemukan baris dengan nilai kosong pada kolom manapun, baris tersebut dihapus menggunakan fungsi `data.dropna()`. Pada dataset Sambiresik, ditemukan 1 (satu) baris dengan missing value yang selanjutnya dihapus, sehingga jumlah data yang dianalisis menjadi 2.817 baris (sebelum pemilihan fitur).

Kedua, pemeriksaan dan penanganan data duplikat. Fungsi `data.duplicated().sum()` digunakan untuk mendeteksi baris yang identik secara keseluruhan dengan baris lainnya. Jika ditemukan duplikat, fungsi `data.drop_duplicates()` digunakan untuk menghapus baris berulang dan hanya mempertahankan satu salinan unik. Pada dataset yang digunakan, tidak ditemukan data duplikat.

Ketiga, verifikasi tipe data. Fungsi `data.dtypes` digunakan untuk memeriksa tipe data setiap kolom. Kolom LUAS bertipe numerik (float64/int64) dan dapat langsung digunakan, sedangkan kolom RTRW bertipe object (string) sehingga memerlukan encoding sebelum diproses oleh K-Means.

Keempat, analisis statistik deskriptif. Fungsi `data.describe()` digunakan untuk mendapatkan ringkasan statistik dataset, meliputi nilai minimum,

maksimum, rata-rata, dan standar deviasi setiap kolom numerik. Dari analisis ini diketahui bahwa nilai LUAS berkisar antara 7 m² (minimum) hingga 65.600 m² (maksimum) dengan rata-rata sekitar 379 m², menunjukkan distribusi yang sangat skewed ke kanan (right-skewed) dan terdapat beberapa outlier dengan luas sangat besar.

3.1.4 Transformasi Data

Transformasi data dilakukan untuk mengubah representasi data dari format aslinya menjadi format yang dapat diterima oleh algoritma K-Means. Dua jenis transformasi utama dilakukan dalam penelitian ini:

Pertama, Label Encoding pada kolom RTRW. Label Encoding adalah teknik encoding kategorikal yang mengubah setiap kategori unik dalam suatu kolom menjadi nilai integer berurutan. Implementasi dilakukan menggunakan kelas `LabelEncoder` dari modul `sklearn.preprocessing`. Proses encoding dilakukan dalam dua tahap: (a) melatih encoder dengan memanggil `label_encoder.fit_transform(data_cluster['RTRW'])`, di mana encoder secara otomatis mengurutkan kategori secara alfabetis dan mengassign nilai integer mulai dari 0; (b) menyimpan hasil encoding dalam kolom baru bernama `RTRW_encoded`. Hasil encoding menunjukkan bahwa 'Kawasan Permukiman Perkotaan' dikodekan sebagai nilai 0 dan 'Kawasan Tanaman Pangan' dikodekan sebagai nilai 1.

Tabel 3.2 Mapping Label Encoding Kolom RTRW

No.	Kategori RTRW (Asli)	Nilai Encoded
1	Kawasan Permukiman Perkotaan	0
2	Kawasan Tanaman Pangan	1

Kedua, seleksi fitur untuk clustering. Setelah proses encoding, dua kolom yang digunakan sebagai fitur input K-Means dipilih, yaitu LUAS dan RTRW_encoded. Kolom NIB tidak diikutsertakan karena berfungsi sebagai identifier unik, sedangkan kolom TAHUN dan LSD tidak digunakan sebagai fitur utama karena tidak relevan dengan tujuan clustering pemanfaatan lahan.

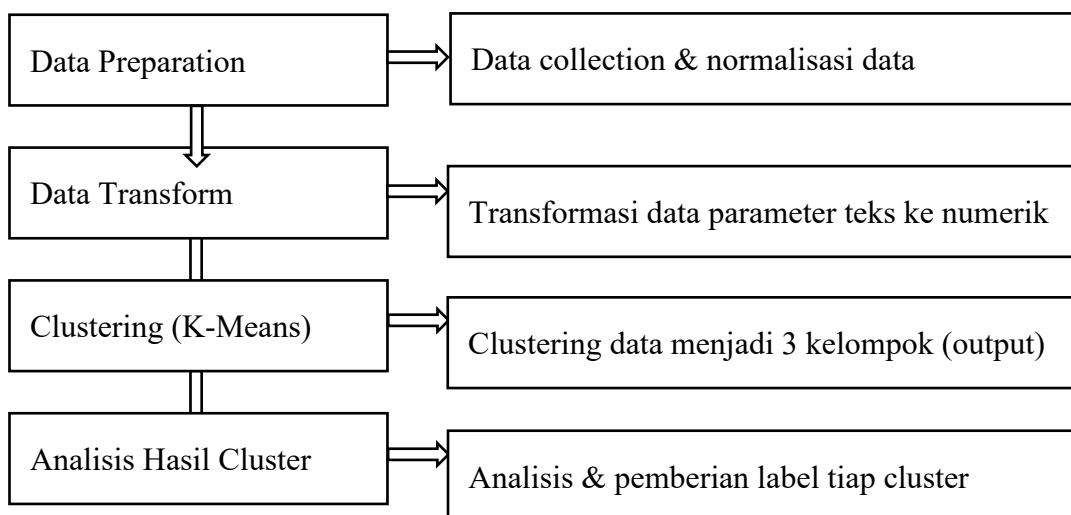
3.4 Desain Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif eksploratif dengan desain komparasi dua model K-Means. Secara garis besar, alur penelitian terdiri atas lima tahap utama yang berurutan: (1) Pengumpulan dan pembersihan data, (2) Transformasi data, (3) Implementasi dua model K-Means (Standar dan Modifikasi), (4) Evaluasi dengan Silhouette Score, dan (5) Interpretasi dan pelabelan klaster. Kedua model diimplementasikan secara terpisah dengan variasi jumlah klaster ($k=2$ sampai $k=5$), kemudian hasilnya dibandingkan berdasarkan nilai Silhouette Score dan distribusi anggota klaster.

Perbedaan mendasar antara dua model yang diimplementasikan adalah sebagai berikut: Model K-Means Standar membiarkan algoritma secara alami membentuk k klaster dari seluruh data tanpa pembatasan fitur kategorikal, sehingga

data RTRW dikodekan sebagai 0 dan 1 dan diperlakukan setara dengan fitur LUAS dalam jarak Euclidean. Model K-Means Modifikasi, di sisi lain, menerapkan constraint berbasis domain pengetahuan: seluruh data Kawasan Tanaman Pangan (RTRW_encoded=1) secara paksa dikelompokkan ke dalam satu kluster tersendiri (kluster ke-k), sementara data Kawasan Permukiman Perkotaan (RTRW_encoded=0) dibagi ke dalam k-1 kluster.

Berikut desain penelitian yang akan dilaksanakan :



Data Preparation : Data preparation berisi langkah-langkah data didapatkan hingga proses normalisasi data yang menjadikan data dapat diolah sesuai dengan kebutuhan peneliti.

Data Transform : Transformasi data dilakukan pada data status RTRW dan LSD. Hal ini dikarenakan kedua data tersebut memiliki nilai berupa teks. Jadi perlu dilakukan transformasi data dari teks menjadi numerik karena klusterisasi K-Means merupakan algoritma yang hanya dapat dijalankan ketika data yang diproses berupa numerik untuk membagi kelompok klasternya.

Clustering (K-Means) : K-Means Clustering berfungsi untuk mengelompokkan data yang telah melalui data preparation dan transform data yaitu normalisasi dan transformasi.

Analisis Hasil Cluster : Setelah klasterisasi selesai, maka langkah terakhir ialah menentukan label klaster yang akan disematkan sesuai dengan pembagian kelompoknya.

3.5 Clustering (K-Means)

Penerapan metode K-Means Clustering yang disajikan mencakup landasan konseptual matematis dari algoritma tersebut, prosedur iteratif yang dilakukan dalam pembentukan klaster, hingga parameter dan skenario implementasi pada dataset pemanfaatan lahan. Pemaparan ini bertujuan untuk memberikan justifikasi analitis secara menyeluruh mengenai bagaimana algoritma K-Means beroperasi mempartisi data guna mencapai tingkat pengelompokan yang paling optimal berdasarkan kedekatan karakteristiknya.

3.1.5 Konsep K-Means

Algoritma K-Means merupakan algoritma clustering berbasis partisi (partitional clustering) yang bekerja dengan cara mempartisi dataset $X = \{x_1, x_2, \dots, x_n\}$ menjadi K kelompok non-overlapping (klaster) $\{C_1, C_2, \dots, C_K\}$ sedemikian sehingga setiap data point x_i berada di tepat satu klaster. Objektif optimisasi K-Means adalah meminimalkan Within-Cluster Sum of Squares (WCSS), yang didefinisikan sebagai:

$$WCSS = \sum_{i=1}^K \sum_{x \in C_i} \|x - \mu_i\|^2$$

di mana μ_i merepresentasikan centroid (titik pusat) dari klaster ke- i , dan $\|\cdot\|^2$ merupakan jarak Euclidean kuadrat. Nilai WCSS yang rendah menandakan klaster yang kompak, di mana anggota klaster berada dekat dengan centroidnya. Pada implementasi ini, jarak Euclidean antara dua titik data $p=(p_1,p_2)$ dan $q=(q_1,q_2)$ dihitung sebagai: $d(p,q) = \sqrt{(p_1-q_1)^2 + (p_2-q_2)^2}$.

Parameter `random_state=42` digunakan dalam inisialisasi KMeans untuk menjamin reproducibility hasil, sehingga setiap eksekusi kode menghasilkan hasil clustering yang identik. Nilai 42 merupakan konvensi umum dalam komunitas machine learning sebagai seed acak yang deterministik.

3.1.6 Tahapan Clustering

Algoritma K-Means diimplementasikan menggunakan kelas KMeans dari library scikit-learn. Berikut adalah penjelasan teknis lengkap setiap tahapan:

Tahap 1: Inisialisasi Centroid. Scikit-learn secara default menggunakan metode inisialisasi 'k-means++' (bukan inisialisasi acak murni) untuk meningkatkan kualitas konvergensi. Metode ini memilih centroid pertama secara acak dari data, lalu centroid berikutnya dipilih dengan probabilitas proporsional terhadap kuadrat jaraknya dari centroid yang sudah ada. Proses ini menghasilkan inisialisasi yang lebih tersebar, mengurangi risiko konvergensi ke local minimum.

Tahap 2: Penugasan Klaster. Setiap data point x_i ditugaskan ke klaster C_j yang centroidnya paling dekat, diformulasikan sebagai: $\text{argmin}_j d(x_i, \mu_j)$, di mana $j \in \{1, 2, \dots, K\}$.

Tahap 3: Pembaruan Centroid. Setelah semua data point ditugaskan, centroid setiap klaster diperbarui sebagai rata-rata aritmetis dari semua anggota klaster: $\mu_i = (1/|C_i|) \sum_{x \in C_i} x$.

Tahap 4: Iterasi dan Konvergensi. Tahap 2 dan 3 diulangi hingga kondisi konvergensi terpenuhi, yaitu tidak ada perubahan penugasan klaster antara iterasi berturutan, atau WCSS tidak berubah secara signifikan. Parameter `max_iter` (default=300) membatasi jumlah maksimum iterasi untuk mencegah loop tak terhingga. Parameter `n_init` (default=10) berarti algoritma dijalankan 10 kali dengan inisialisasi berbeda, dan hasil terbaik (WCSS terendah) yang dipilih.

3.1.7 Implementasi Model K-Means Standar

Pada model K-Means Standar, fungsi `perform_clustering(data, k)` mengimplementasikan klasterisasi standar yang memproses seluruh dataset tanpa pembatasan berbasis kategori. Pseudocode dari fungsi ini adalah sebagai berikut:

```
def perform_clustering(data, k):
    X = data[['LUAS', 'RTRW_encoded']]
    kmeans = KMeans(n_clusters=k, random_state=42)
    data['Cluster'] = kmeans.fit_predict(X)
    # Menampilkan statistik per klaster
    for cluster_num in range(k):
        cluster_data = data[data['Cluster'] == cluster_num]
        print(f'Klaster {cluster_num}: {len(cluster_data)} data')
        print(f'Luas Min={cluster_data["LUAS"].min()},
Max={cluster_data["LUAS"].max()}')
```

Model ini memperlakukan nilai RTRW_encoded (0 atau 1) dan nilai LUAS sebagai dua dimensi yang setara dalam ruang fitur dua dimensi. Karena skala LUAS (7 hingga 65.600 m²) jauh lebih besar dibandingkan RTRW_encoded (0 atau 1), algoritma K-Means cenderung menghasilkan klaster yang didominasi oleh perbedaan luasan, sementara perbedaan kategori RTRW memiliki pengaruh yang relatif kecil terhadap pembentukan.

3.1.8 Implementasi Model K-Means Modifikasi

Model K-Means Modifikasi dirancang berdasarkan pengetahuan domain (domain knowledge) bahwa Kawasan Tanaman Pangan memiliki karakteristik penggunaan lahan yang secara fundamental berbeda dengan Kawasan Permukiman Perkotaan. Oleh karena itu, pendekatan modifikasi menerapkan constraint hard assignment, di mana seluruh data Kawasan Tanaman Pangan secara paksa ditempatkan dalam satu klaster tersendiri, sementara K-Means bebas mengklasterkan data Kawasan Permukiman Perkotaan ke dalam k-1 klaster.

```
def clustering_kmeans(data, k):
    pemukiman_data = data[data['RTRW_encoded'] == 0]
    tanaman_pangan_data = data[data['RTRW_encoded'] == 1]
    X_pemukiman = pemukiman_data[['LUAS', 'RTRW_encoded']]
    kmeans_pemukiman = KMeans(n_clusters=k-1, random_state=42)
    pemukiman_data['Cluster'] =
    kmeans_pemukiman.fit_predict(X_pemukiman)
    tanaman_pangan_data['Cluster'] = k - 1 # klaster terakhir
    final_data = pd.concat([pemukiman_data, tanaman_pangan_data])
```


Dengan pendekatan ini, ketika $k=4$, data Kawasan Permukiman Perkotaan diklasterkan menjadi 3 sub-kelompok (Klaster 0, 1, 2) berdasarkan variasi luasan, sedangkan data Kawasan Tanaman Pangan seluruhnya masuk ke Klaster 3. Pendekatan ini secara eksplisit memastikan bahwa zonasi RTRW dihormati sebagai batasan klaster yang tidak dapat dilangkahi oleh optimisasi K-Means.

3.1.9 Analisis Hasil Cluster

Setelah proses clustering selesai, analisis dilakukan untuk setiap klaster yang terbentuk. Analisis mencakup tiga aspek utama:

(1) Distribusi anggota: menghitung jumlah total bidang tanah dalam setiap klaster menggunakan `value_counts()` pada kolom Cluster. Distribusi yang sangat tidak seimbang (misalnya satu klaster dengan 90% data) dapat mengindikasikan bahwa nilai k belum optimal.

(2) Statistik luas: menentukan nilai minimum dan maksimum luas bidang tanah dalam setiap klaster untuk memahami rentang luas yang merepresentasikan masing-masing kelompok.

(3) Distribusi RTRW per klaster: mengidentifikasi proporsi data Kawasan Permukiman Perkotaan (`RTRW_encoded=0`) dan Kawasan Tanaman Pangan (`RTRW_encoded=1`) dalam setiap klaster, yang menginformasikan seberapa murni klaster tersebut secara kategorikal.

Evaluasi kualitas klaster menggunakan Silhouette Score, sebuah metrik yang mengukur seberapa baik setiap data point berada dalam klasternya sendiri dibandingkan dengan klaster terdekat lainnya. Nilai Silhouette Score berkisar antara -1 hingga +1, di mana nilai mendekati +1 menandakan klaster yang sangat

terdefinisi dengan baik, nilai 0 menandakan tumpang tindih antar klaster, dan nilai negatif menandakan kemungkinan kesalahan penugasan data. Formula Silhouette Score untuk satu data point x_i adalah: $s(i) = (b(i) - a(i)) / \max\{a(i), b(i)\}$, di mana $a(i)$ adalah rata-rata jarak intra-klaster (jarak rata-rata dari x_i ke semua anggota klasternya) dan $b(i)$ adalah rata-rata jarak inter-klaster terkecil (rata-rata jarak dari x_i ke semua anggota klaster terdekat yang bukan klasternya).

BAB IV

HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan analisis dari seluruh tahapan yang telah dirancang pada Bab III. Pembahasan mencakup deskripsi data, hasil preprocessing, implementasi dua model K-Means (Standar dan Modifikasi) dengan empat variasi nilai k , evaluasi komparatif menggunakan Silhouette Score, serta interpretasi hasil clustering dalam konteks pemanfaatan lahan di Desa Sambiresik

4.1 Deskripsi Data

Dataset yang digunakan dalam penelitian ini merupakan data pertanahan resmi Desa Sambiresik yang diperoleh dari BPN Kabupaten Kediri. Dataset memiliki total 2.818 baris data (sebelum preprocessing), masing-masing merepresentasikan satu bidang tanah yang terdaftar. Dataset memiliki lima kolom atribut: NIB (Nomor Identifikasi Bidang Tanah), TAHUN (tahun pendaftaran), LUAS (luas bidang tanah dalam m^2), RTRW (zonasi tata ruang), dan LSD (status lahan sawah dilindungi).

Kolom NIB memuat nilai unik dari 1781 hingga 4598, berfungsi sebagai pengenal bidang tanah. Kolom TAHUN menunjukkan bahwa sebagian besar data tidak memiliki tahun terdaftar (nilai 0, sebanyak 1.463 data) sementara 1.355 data tercatat pada tahun 2023. Kolom LUAS mencakup rentang yang sangat lebar dari $7 m^2$ hingga $65.600 m^2$ (nilai outlier dari NIB 2173), dengan median sekitar $192 m^2$. Kolom RTRW terbagi menjadi 'Kawasan Permukiman Perkotaan' (2.230 bidang,

79,13%) dan 'Kawasan Tanaman Pangan' (588 bidang, 20,87%). Kolom LSD memiliki dua nilai: 'Terbuka' (2.379 bidang) dan 'LSD' (437 bidang).

Untuk keperluan klasterisasi, hanya kolom LUAS dan RTRW yang digunakan sebagai fitur, berdasarkan pertimbangan relevansi domain bahwa kedua atribut tersebut paling langsung merepresentasikan karakteristik pemanfaatan lahan yang ingin dikelompokkan.

Tabel 4.1 Statistik Deskriptif Dataset

Statistik	NIB	TAHUN	LUAS (m ²)	RTRW
Count	2.818	2.818	2.818	2.818
Min	1781	0	7	-
Max	4598	2023	65.600	-
Mean	3189.5	870.0	~379	-
Unique	2818	2 nilai	Beragam	2 kategori

Sumber: Data BPN Kabupaten Kediri, diolah, 2025

4.2 Preprocessing Data

Tahap preprocessing diimplementasikan secara sistematis menggunakan kode Python pada Google Colab. Berikut adalah penjelasan detail setiap langkah yang dilakukan beserta hasilnya:

Langkah pertama adalah pembacaan data. File CSV Sambiresik.csv dibaca menggunakan `pd.read_csv('/content/Sambiresik.csv')` yang memuat seluruh 2.818 baris data ke dalam DataFrame Pandas. Verifikasi pembacaan dilakukan dengan

menampilkan beberapa baris pertama menggunakan perintah `data.head()` dan memeriksa dimensi dengan `data.shape`.

Langkah kedua adalah pendeteksian dan penanganan missing values. Pemeriksaan dilakukan dengan `data.isnull().sum()` yang mengembalikan jumlah nilai kosong per kolom. Hasil menunjukkan terdapat 1 (satu) baris dengan nilai NaN pada kolom tertentu. Baris tersebut dihapus menggunakan `data.dropna()`, menghasilkan dataset bersih dengan 2.817 baris. Keputusan untuk menghapus (bukan mengisi) baris dengan missing value didasarkan pada pertimbangan bahwa satu data point tidak akan mempengaruhi signifikansi hasil dengan dataset lebih dari 2.800 baris.

Langkah ketiga adalah pendeteksian duplikat. Pemeriksaan `data.duplicated().sum()` mengembalikan nilai 0 (nol), mengkonfirmasi bahwa tidak ada baris yang identik dalam dataset. Ini menunjukkan bahwa setiap NIB merepresentasikan bidang tanah yang benar-benar berbeda.

Langkah keempat adalah Label Encoding pada kolom RTRW. `LabelEncoder` dari `sklearn.preprocessing` digunakan untuk mengkonversi nilai string menjadi integer. Setelah `fit_transform()` diterapkan, mapping encoding terbentuk sebagai berikut: 'Kawasan Permukiman Perkotaan' $\rightarrow$ 0, dan 'Kawasan Tanaman Pangan' $\rightarrow$ 1 (pengurutan alfabetis otomatis). Kolom baru `RTRW_encoded` ditambahkan ke `DataFrame`.

Tabel 4.2 Contoh Data Setelah Proses Encoding

NIB	RTRW (Asli)	RTRW_encoded
1781	Kawasan Permukiman Perkotaan	0
4594	Kawasan Tanaman Pangan	1
4595	Kawasan Tanaman Pangan	1
4598	Kawasan Permukiman Perkotaan	0

Langkah kelima adalah seleksi fitur dan persiapan data untuk clustering. Dua kolom dipilih: LUAS dan RTRW_encoded. Kolom RTRW asli dihapus dari DataFrame clustering (`data_cluster.drop(columns=['RTRW'])`) karena telah digantikan oleh representasi numeriknya. DataFrame akhir yang siap untuk clustering memiliki dua kolom: LUAS (float/int) dan RTRW_encoded (int, bernilai 0 atau 1).

Langkah keenam adalah visualisasi eksploratori. Sebelum clustering, dilakukan tiga visualisasi: (1) Histogram distribusi LUAS menggunakan `sns.histplot()` untuk memahami sebaran data luasan, yang menunjukkan distribusi sangat miring ke kanan (right-skewed) dengan mayoritas bidang tanah berukuran kecil ($<500 \text{ m}^2$);

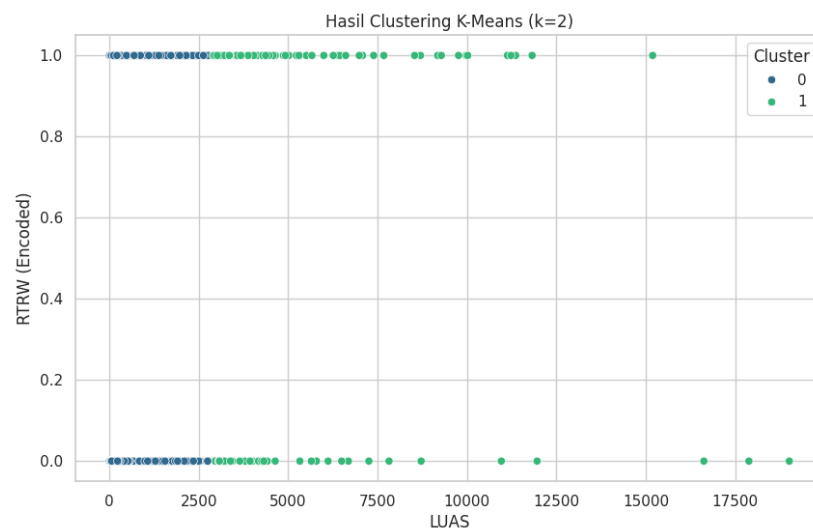
(2) Scatter plot LUAS vs NIB untuk melihat sebaran data terhadap nomor identifikasi; (3) Scatter plot RTRW vs LUAS untuk memahami distribusi luasan per kategori RTRW, yang mengkonfirmasi bahwa Kawasan Tanaman Pangan umumnya memiliki luas yang lebih bervariasi dibandingkan Kawasan Permukiman Perkotaan.

4.3 Clustering K-Means

Proses clustering diimplementasikan menggunakan dua model yang berbeda. Fungsi `perform_clustering()` mengimplementasikan Model K-Means Standar (tanpa constraint RTRW), sedangkan fungsi `clustering_kmeans()` mengimplementasikan Model K-Means Modifikasi (dengan constraint RTRW). Kedua model dijalankan untuk nilai $k=2$, $k=3$, $k=4$, dan $k=5$.

4.3.1 Pemilihan K dan Hasil Clustering pada Model K -Means Standar

1. $K=2$



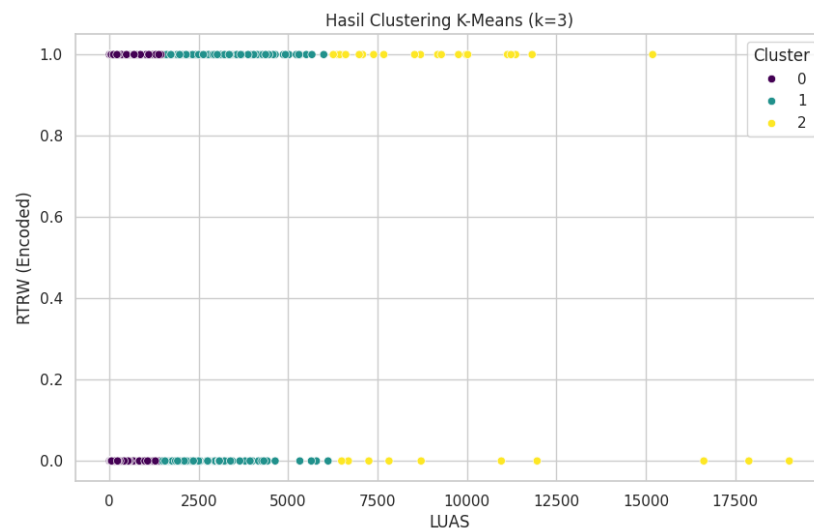
Gambar 4.3.1 Pemilihan $K=2$

Hasil dari bentuk Clustering Model K-Means Standar dengan $K=2$ ditunjukkan pada

Tabel 4.3.1 $K=2$

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
0	2.690	7,0	2.804,0	2.187	503
1	126	2.892,0	19.010,0	43	83

2. $K=3$

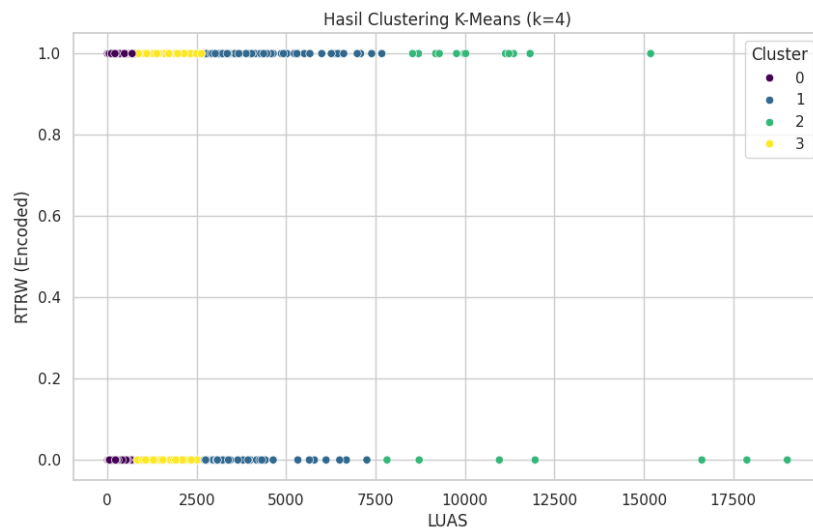
Gambar 4.3.2 Pemilihan $K=3$

Hasil dari bentuk Clustering Model K-Means Standar dengan $K=3$ ditunjukkan pada

Tabel 4.3.2 $K=3$

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
0	2.523	7,0	1.465,0	2.121	402
1	260	1.478,0	6.117,0	99	161
2	33	6.263,0	19.010,0	10	23

3. K=4



Gambar 4.3.3 Pemilihan K=4

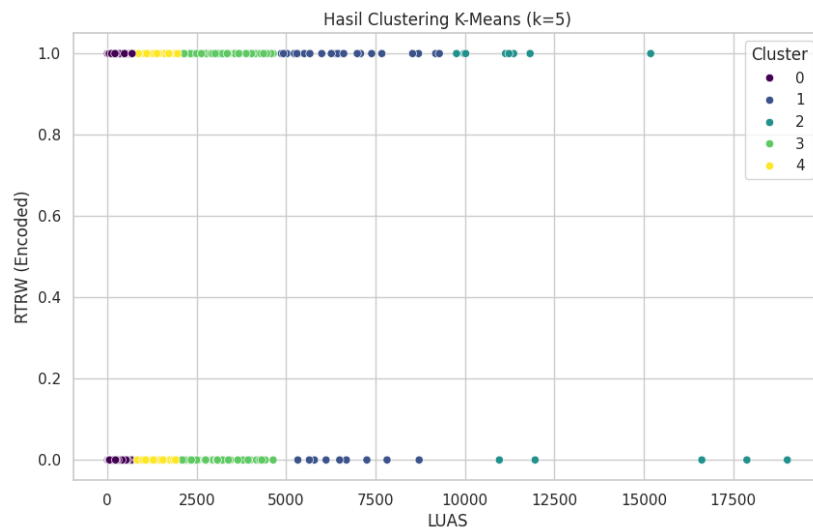
Hasil dari bentuk Clustering Model K-Means Standar dengan K=4 ditunjukkan pada

Tabel 4.3.3 Pemilihan K=4

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
0	2.244	7,0	798,0	1.980	264

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
1	118	2.718,0	7.675,0	40	78
2	20	7.821,0	19.010,0	7	13
3	434	800,0	2.667,0	203	231

4. K=5



Gambar 4.3.4 Pemilihan K=5

Hasil dari bentuk Clustering Model K-Means Standar dengan K=5 ditunjukkan pada

Tabel 4.3.4 Pemilihan K=5

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
0	2.193	7,0	700,0	1.944	249

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
1	35	4.865,0	9.283,0	9	26
2	14	9.763,0	19.010,0	5	9
3	139	2.098,0	4.640,0	56	83
4	435	702,0	2.095,0	216	219

Evaluasi untuk Model K-Means Standar dengan percobaan K dari 2 hingga 5 ditunjukkan pada

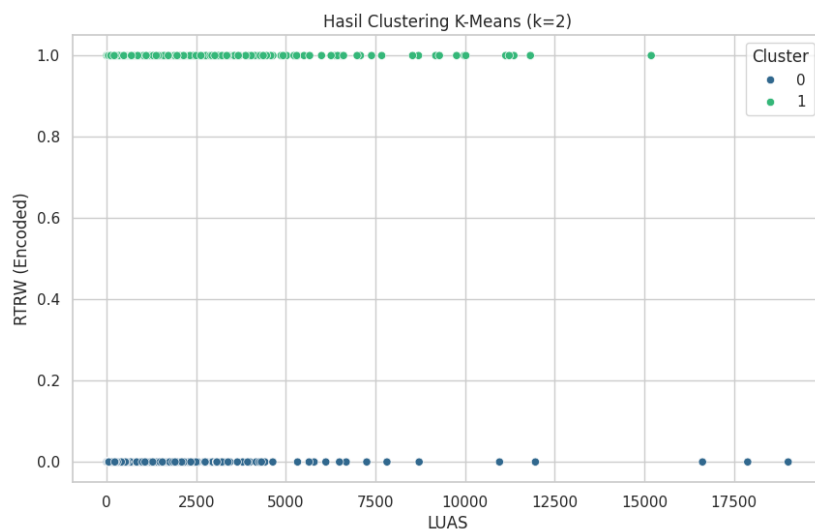
Tabel 4.3.5 Silhouette Score I

K	Jumlah Data Cluster	Silhouette Score
2	Cluster 0 = 2.193	0.7575
	Cluster 1 = 435	
3	Cluster 0 = 2.193	0.7575
	Cluster 1 = 435	
	Cluster 2 = 139	
4	Cluster 0 = 2.193	0.7575
	Cluster 1 = 435	
	Cluster 2 = 139	
	Cluster 3 = 35	
5	Cluster 0 = 2.193	0.7575
	Cluster 1 = 435	

K	Jumlah Data Cluster	Silhouette Score
	Cluster 2 = 139	
	Cluster 3 = 35	
	Cluster 4 = 14	

4.3.2 Pemilihan K dan Hasil *Clustering* pada Model *K-Means* Modifikasi

1. $K = 2$



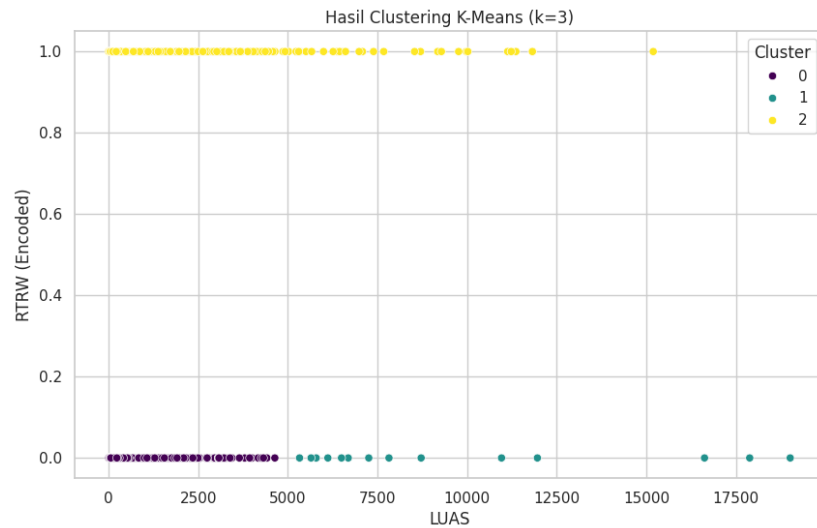
Gambar 4.4.1 Pemilihan $K=2$ Mod

Hasil dari bentuk Clustering Model *K-Means* Modifikasi dengan $K=2$ ditunjukkan pada

Tabel 4.4.1 Pemilihan $K=2$ Mod

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
0	2.230	10,0	19.010,0	2.230	0
1	586	7,0	15.190,0	0	586

2. $K = 3$



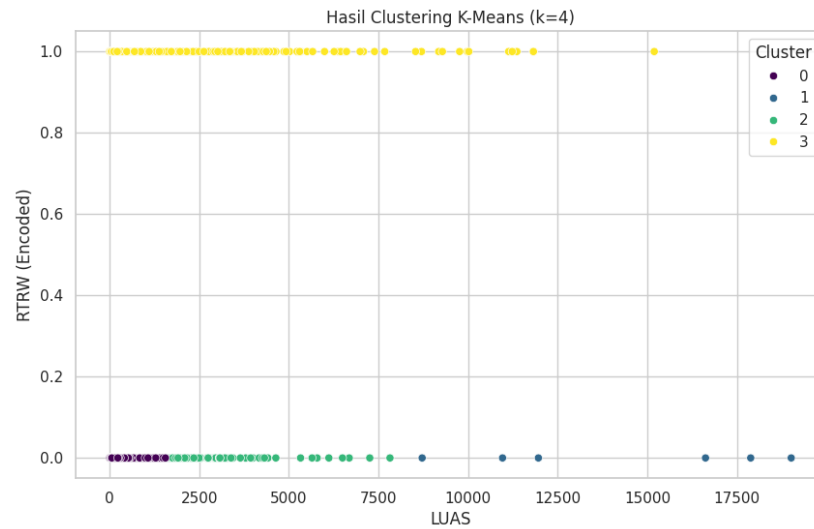
Gambar 4.4.2 Pemilihan $K=3$ Mod

Hasil dari bentuk Clustering Model K-Means Modifikasi dengan $K=3$ ditunjukkan pada

Tabel 4.4.2 Pemilihan $K=3$ Mod

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
0	2.216	10,0	4.638,0	2.216	0
1	14	5.328,0	19.010,0	14	0
2	586	7,0	15.190,0	0	586

3. $K = 4$



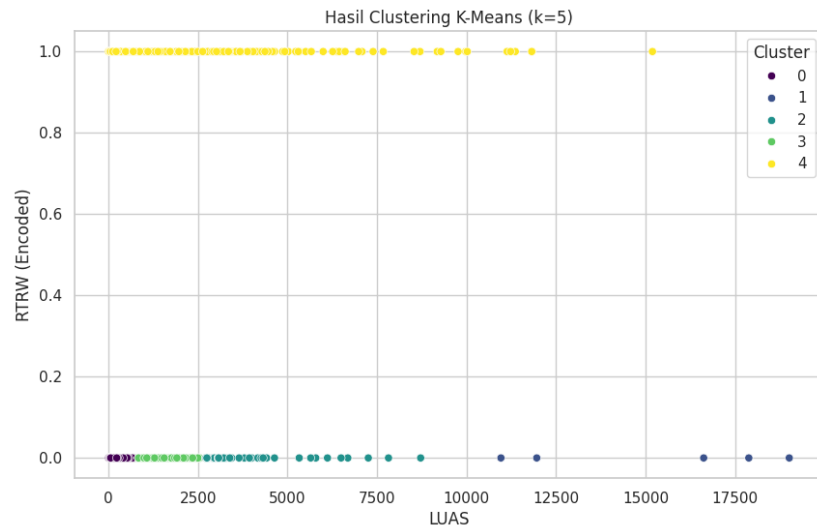
Gambar 4.4.3 Pemilihan K=4 Mod

Hasil dari bentuk Clustering Model K-Means Modifikasi dengan K=4 ditunjukkan pada

Tabel 4.4.3 Pemilihan K=4 Mod

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
0	2.135	10,0	1.564,0	2.135	0
1	6	8.718,0	19.010,0	6	0
2	89	1.575,0	7.821,0	89	0
3	586	7,0	15.190,0	0	586

4. K = 5



Gambar 4.4.4 Pemilihan K=5 Mod

Hasil dari bentuk Clustering Model K-Means Modifikasi dengan K=5 ditunjukkan pada

Tabel 4.4.4 Pemilihan K=5 Mod

Cluster	Jumlah Data	Luas		RTRW	
		Min (m ²)	Max (m ²)	0	1
0	1.953	10,0	721,0	1.953	0
1	5	10.960,0	19.010,0	5	0
2	42	2.718,0	8.718,0	42	0
3	230	724,0	2.548,0	230	0
4	586	7,0	15.190,0	0	586

Evaluasi untuk Model K-Means Modifikasi dengan percobaan K dari 2 hingga 5 ditunjukkan pada

Tabel 4.3.5 Silhouette Score II

K	Jumlah data cluster	Silhouoette Score
2	Cluster 0 = 2.230	0.4632
	Cluster 1 = 586	
3	Cluster 0 = 2.216	0,4859
	Cluster 1 = 14	
	Cluster 2 = 586	
4	Cluster 0 = 2.135	0,5098
	Cluster 1 = 6	
	Cluster 2 = 89	
	Cluster 3 = 586	
5	Cluster 0 = 1.953	0,4984
	Cluster 1 = 5	
	Cluster 2 = 42	
	Cluster 3 = 230	
	Cluster 4 = 586	

4.3.3 Perbandingan Model K-Means

Tabel 4.5.1 Perbandingan

K	Cluster dan Jumlah Data		Silhouette Score	
	Standar	Modifikasi	Standar	Modifikasi
2	Cluster 0 = 2.193	Cluster 0 = 2.230	0.7575	0.4632
	Cluster 1 = 435	Cluster 1 = 586		

K	Cluster dan Jumlah Data		Silhouette Score	
	Standar	Modifikasi	Standar	Modifikasi
3	Cluster 0 = 2.193	Cluster 0 = 2.216	0.7575	0,4859
	Cluster 1 = 435	Cluster 1 = 14		
	Cluster 2 = 139	Cluster 2 = 586		
4	Cluster 0 = 2.193	Cluster 0 = 2.135	0.7575	0,5098
	Cluster 1 = 435	Cluster 1 = 6		
	Cluster 2 = 139	Cluster 2 = 89		
	Cluster 3 = 35	Cluster 3 = 586		
5	Cluster 0 = 2.193	Cluster 0 = 1.953	0.7575	0,4984
	Cluster 1 = 435	Cluster 1 = 5		
	Cluster 2 = 139	Cluster 2 = 42		
	Cluster 3 = 35	Cluster 3 = 230		
	Cluster 4 = 14	Cluster 4 = 586		

Berdasarkan hasil perbandingan, K-Means Standar menunjukkan Silhouette Score yang konstan di 0.7575 untuk semua nilai k (2 hingga 5). Hal ini menunjukkan bahwa clustering tetap stabil, tetapi untuk menentukan nilai k yang optimal, keseimbangan antara pemisahan yang baik dan jumlah klaster yang tidak terlalu banyak perlu dipertimbangkan. Karena nilai Silhouette Score tidak berubah, k=2 dapat dianggap sebagai pilihan terbaik karena lebih sederhana dengan jumlah klaster yang minimal, namun tetap memberikan pemisahan yang baik.

Sementara itu, pada Modifikasi Model K-Means, Silhouette Score meningkat dari $k=2$ hingga $k=4$, tetapi kembali turun pada $k=5$. Skor tertinggi diperoleh pada $k=4$ dengan 0.5098, yang menunjukkan bahwa pemisahan kluster paling optimal terjadi pada nilai ini. Secara rinci, Silhouette Score pada $k=2$ adalah 0.4632, meningkat menjadi 0.4859 pada $k=3$, mencapai puncaknya di 0.5098 pada $k=4$, namun turun kembali ke 0.4984 pada $k=5$. Oleh karena itu, dalam metode modifikasi ini, $k=4$ adalah nilai optimal karena memberikan kualitas kluster yang lebih baik dibandingkan k lainnya.

Kesimpulannya, K-Means Standar optimal pada $k=2$, karena memberikan pemisahan yang baik dengan kompleksitas rendah, sementara Modifikasi Model K-Means optimal pada $k=4$, karena pada nilai ini diperoleh Silhouette Score tertinggi, menunjukkan pemisahan kluster yang lebih optimal. Jika pemisahan kluster yang lebih kuat dan lebih jelas diutamakan, maka K-Means Standar dengan $k=2$ lebih disarankan, namun jika ingin eksplorasi lebih lanjut dengan model modifikasi, maka $k=4$ pada Modifikasi Model K-Means menjadi pilihan yang lebih baik.

4.4 Interpretasi Hasil dengan Topik

Hasil klastering menggunakan metode K-Means dengan berbagai nilai k memberikan wawasan penting dalam analisis pemanfaatan lahan berdasarkan luas dan kategori rencana tata ruang wilayah. Pada penelitian ini, dua model K-Means diterapkan, yaitu Model K-Means Standar dan Model K-Means Modifikasi, guna menentukan pemisahan kluster yang optimal.

Pada Model K-Means Standar, dengan $k = 2$, data terbagi menjadi dua kelompok utama, di mana klaster 0 mencakup sebagian besar data dan klaster 1 berisi data dengan luas yang lebih besar. Peningkatan jumlah klaster hingga $k = 5$ menunjukkan pemisahan yang lebih rinci, namun nilai Silhouette Score tetap stabil pada angka 0,7575 untuk semua nilai k , yang mengindikasikan bahwa pemisahan tetap konsisten dan stabil tanpa peningkatan signifikan dalam kualitas klaster. Oleh karena itu, nilai $k = 2$ dipilih sebagai yang optimal karena menghasilkan pemisahan yang baik dengan kompleksitas minimal.

Sementara itu, pada Model K-Means Modifikasi, "Kawasan Tanaman Pangan" dikelompokkan ke dalam klaster terakhir, sementara "Kawasan Permukiman Perkotaan" dikategorikan berdasarkan luasnya. Hasil evaluasi menunjukkan bahwa nilai Silhouette Score meningkat dari $k = 2$ (0,4632) ke $k = 4$ (0,5098), namun mengalami penurunan kembali pada $k = 5$ (0,4984). Oleh karena itu, nilai optimal untuk model ini ditetapkan pada $k = 4$.

Berdasarkan hasil perbandingan, Model K-Means Standar menunjukkan performa optimal pada $k = 2$ karena menghasilkan pemisahan yang baik dengan kompleksitas rendah, sedangkan Model K-Means Modifikasi menunjukkan hasil optimal pada $k = 4$ dengan nilai Silhouette Score tertinggi (0,5098), yang mengindikasikan pemisahan klaster yang lebih optimal. Kesimpulannya, jika prioritas diberikan pada pemisahan yang lebih stabil dan sederhana, maka Model K-Means Standar dengan $k = 2$ merupakan pilihan yang lebih tepat. Namun, jika eksplorasi lebih lanjut dengan pendekatan modifikasi diperlukan, maka Model K-Means Modifikasi dengan $k = 4$ menjadi opsi yang lebih baik.

Berdasarkan hasil penerapan metode K-Means Clustering, bidang tanah di wilayah penelitian terbagi menjadi beberapa kelompok atau klaster. Pembentukan klaster ini mempertimbangkan luasan bidang tanah serta fungsi pemanfaatannya yang mengacu pada data Rencana Tata Ruang Wilayah (RTRW) dan status Lahan Sawah Dilindungi (LSD). Selain itu, klasifikasi fungsi pemukiman juga merujuk pada ketentuan Peraturan Menteri Perumahan Rakyat Republik Indonesia Nomor 10 Tahun 2012 tentang penyelenggaraan perumahan dan kawasan permukiman dengan hunian berimbang, serta kebijakan teknis dari Kementerian PUPR dan ATR/BPN Kabupaten Kediri.

1. Cluster Pemukiman Kecil sebagai Rumah Hunian Tunggal

Cluster ini mencakup bidang tanah dengan luasan relatif kecil, berkisar antara 60 m² hingga 300 m². Karakteristik bidang tanah dalam cluster ini umumnya digunakan untuk rumah hunian tunggal yang berdiri di atas lahan mandiri milik perorangan. Berdasarkan kajian Suryo (2017), rumah hunian sederhana membutuhkan lahan minimal 60 m², sehingga bidang tanah dalam cluster ini ideal untuk hunian skala rumah tangga dengan model rumah tapak atau rumah tinggal tunggal.

2. Cluster Pemukiman Sedang Kawasan sebagai Perumahan Mewah (Non KPR)

Cluster kedua mewakili bidang tanah dengan luasan antara 1.000 m² hingga 5.000 m². Bidang tanah dengan luasan ini umumnya telah berkembang menjadi kawasan perumahan eksklusif, seperti perumahan mewah non-subsidi (non-KPR) yang mengutamakan konsep low density housing. Kawasan ini memiliki akses jalan yang lebar (minimal 6 meter sesuai ketentuan Kementerian PUPR)

dan fasilitas lingkungan yang lebih lengkap, seperti taman, fasilitas olahraga, serta sistem keamanan terpadu.

3. Cluster Pemukiman Besar Kawasan sebagai Perumahan Menengah (KPR dan/atau Non KPR)

Cluster ini mencakup bidang tanah dengan luasan berkisar antara 5.000 m² hingga 20.000 m². Karakteristiknya mengarah pada pengembangan kawasan perumahan skala menengah, yang menyasar segmen kelas menengah ke atas maupun penerima fasilitas pembiayaan KPR komersial. Kawasan ini umumnya sudah memiliki site plan terintegrasi dengan infrastruktur pendukung seperti akses jalan utama, saluran air, dan jaringan listrik sesuai standar teknis dari ATR/BPN dan Kementerian PUPR.

4. Cluster Cluster Pemukiman Sangat Besar Kawasan sebagai Perumahan Sederhana (KPR)

Cluster keempat mewakili bidang tanah dengan luasan lebih dari 20.000 m². Luasan ini umumnya diperuntukkan bagi pengembangan kawasan perumahan berskala besar, seperti perumahan rakyat bersubsidi (KPR berskala masal). Kawasan ini dikembangkan dengan konsep hunian berimbang, di mana proporsi hunian sederhana, menengah, dan mewah diatur sesuai ketentuan Permenpera Nomor 10 Tahun 2012. Akses jalan utama memiliki lebar minimal 6 meter, sementara setiap unit rumah memiliki luasan minimal 60 m² dengan total unit tidak kurang dari 10 unit per kawasan. Dalam beberapa kasus, kawasan ini juga dilengkapi dengan fasilitas umum seperti tempat ibadah, taman bermain, dan ruang terbuka hijau.

5. Cluster Kawasan Pertanian dan/atau Perkebunan

Cluster terakhir mencakup bidang tanah dengan luasan yang sangat beragam, mulai dari 500 m² hingga puluhan hektar, tergantung pada pola kepemilikan dan sistem pengelolaan lahan. Cluster ini didominasi oleh lahan sawah, lahan perkebunan, serta lahan produktif lainnya yang statusnya sebagian besar masuk dalam Lahan Sawah Dilindungi (LSD) berdasarkan ketetapan RTRW dan kebijakan ATR/BPN. Karakteristik bidang tanah dalam cluster ini cenderung berada di kawasan perdesaan atau pinggiran kota, dengan infrastruktur dasar yang masih terbatas.

Melalui hasil clustering tersebut, diharapkan dapat diperoleh pemetaan pemanfaatan lahan yang lebih sistematis, serta mendukung pemerintah daerah dalam melakukan monitoring, pengendalian tata ruang, serta perencanaan pembangunan berbasis spasial yang lebih efektif. Pemetaan ini juga dapat menjadi dasar rekomendasi dalam proses penerbitan izin lokasi, pengendalian alih fungsi lahan, hingga penyusunan kebijakan perumahan dan kawasan permukiman yang berkelanjutan.

BAB V

PENUTUP

3.1 Kesimpulan

Penelitian Berdasarkan hasil penelitian yang telah dilakukan, diperoleh kesimpulan sebagai berikut:

1. Metode K-Means Clustering terbukti mampu memenuhi kebutuhan pengelompokan data pemanfaatan lahan yang besar, beragam, dan kompleks di Desa Sambiresik. Dari 2.817 bidang tanah dengan rentang luasan yang sangat heterogen, mulai dari 7 m² hingga lebih dari 65.000 m² dan dua zonasi RTRW yang berbeda, algoritma K-Means berhasil mengklasifikasikan data secara objektif ke dalam klaster-klaster yang bermakna. Model K-Means Modifikasi dengan k=4 menghasilkan pengelompokan paling optimal berdasarkan nilai Silhouette Score tertinggi (0,5098), menghasilkan empat klaster yang mencerminkan karakteristik pemanfaatan lahan secara spesifik: hunian tunggal skala kecil, kawasan perumahan menengah (KPR dan/atau non-KPR), kawasan perumahan berskala besar, serta kawasan pertanian dan perkebunan.
2. Hasil klasterisasi pada Model K-Means Modifikasi dengan k=4 yang dipilih sebagai model paling representatif karena menghormati batas zonasi RTRW secara eksplisit menghasilkan empat klaster dengan karakteristik dan distribusi sebagai berikut:

2.1 Klaster 0, Pemukiman Kecil / Hunian Tunggal (2.135 bidang tanah, 75,8% dari total data Kawasan Permukiman Perkotaan): mencakup bidang tanah dengan rentang luas 10–1.564 m², merepresentasikan hunian tunggal skala rumah tangga yang berdiri di atas lahan mandiri perorangan. Klaster ini merupakan klaster terbesar dan paling dominan dalam dataset, mencerminkan bahwa mayoritas pemanfaatan lahan permukiman di Desa Sambiresik berskala kecil.

2.2 Klaster 2, Pemukiman Menengah / Kawasan Perumahan KPR dan/atau Non-KPR (89 bidang tanah, 3,2% dari total data Kawasan Permukiman Perkotaan): mencakup bidang tanah dengan rentang luas 1.575–7.821 m², yang sesuai untuk pengembangan kawasan perumahan berskala menengah dengan konsep hunian berimbang sesuai Permenpera No. 10 Tahun 2012.

2.3 Klaster 1, Pemukiman Besar / Kawasan Perumahan Mewah Non-KPR (6 bidang tanah, 0,2% dari total data Kawasan Permukiman Perkotaan): mencakup bidang tanah dengan rentang luas 8.718–19.010 m², merepresentasikan lahan berskala besar yang berpotensi dikembangkan sebagai kawasan perumahan eksklusif atau fasilitas umum berskala wilayah.

2.4 Klaster 3, Kawasan Pertanian dan/atau Perkebunan (586 bidang tanah, 20,8% dari total keseluruhan data): mencakup seluruh bidang tanah berzonasi Kawasan Tanaman Pangan dengan rentang luas 7–15.190 m². Sebagian besar bidang tanah dalam klaster ini berstatus Lahan

Sawah Dilindungi (LSD), yang secara hukum tidak dapat dialihfungsikan dan memerlukan pemantauan ketat terhadap tekanan alih fungsi lahan akibat urbanisasi.

3. Hasil klasterisasi yang terbentuk dapat dimanfaatkan sebagai acuan berbasis data yang konkret bagi BPN dan Dinas Perumahan dan Kawasan Permukiman (Perkim). Dari sisi perizinan PBG, profil tiap klaster — mencakup rentang luasan tipikal dan zonasi RTRW yang dominan — memungkinkan verifikasi kesesuaian fungsi lahan dilakukan secara lebih cepat dan terstandar, tanpa harus menilai setiap bidang tanah secara kasuistis. Dari sisi pengendalian alih fungsi lahan, Klaster 3 yang mengidentifikasi 586 bidang tanah pertanian (20,8% dari total data) memberikan dasar data bagi pemerintah daerah untuk memantau dan melindungi lahan produktif dari tekanan urbanisasi yang semakin masif, sekaligus mendukung evaluasi kesesuaian RTRW secara berkala setiap lima tahun.

Hasil penelitian ini memiliki implikasi praktis bagi beberapa pihak. Bagi BPN Kabupaten Kediri, profil klaster yang dihasilkan dapat diintegrasikan ke dalam sistem informasi pertanahan sebagai lapisan klasifikasi tambahan yang memperkaya data administratif yang sudah ada, khususnya untuk mendukung layanan pertanahan berbasis digital sesuai amanat Perpres No. 18 Tahun 2020. Bagi Dinas Perkim, klaster pemanfaatan lahan dapat dijadikan referensi dalam penyusunan standar teknis perizinan PBG berdasarkan tipologi kawasan misalnya, bidang tanah dalam rentang Klaster 0 (10–1.564 m²) cukup diproses dengan

persyaratan hunian tunggal, sedangkan bidang tanah dalam Klaster 2 (1.575–7.821 m²) perlu diverifikasi terhadap ketentuan pengembangan kawasan perumahan. Bagi pemerintah daerah secara umum, pemetaan klaster ini dapat menjadi salah satu masukan dalam proses revisi RTRW berkala, khususnya untuk mengevaluasi apakah kondisi pemanfaatan lahan aktual masih sesuai dengan zonasi yang telah ditetapkan, terutama pada 586 bidang lahan pertanian yang teridentifikasi rentan terhadap tekanan alih fungsi.

3.2 Saran

Berdasarkan temuan dan keterbatasan penelitian ini, disampaikan beberapa saran untuk penelitian maupun kebijakan selanjutnya:

1. Integrasi data spasial (GIS): Hasil klasterisasi dalam penelitian ini bersifat tabular. Penelitian lanjutan disarankan mengintegrasikan koordinat geografis setiap bidang tanah agar klaster dapat divisualisasikan dalam bentuk peta tematik, sehingga lebih mudah diinterpretasikan oleh instansi terkait.
2. Pengayaan variabel: Penambahan variabel seperti harga tanah, aksesibilitas infrastruktur, kepadatan bangunan, atau status kepemilikan lahan diperkirakan akan menghasilkan klaster yang lebih representatif terhadap kondisi pemanfaatan lahan yang sesungguhnya.
3. Perbandingan metode clustering: Penelitian lanjutan disarankan membandingkan hasil K-Means dengan metode clustering lain seperti

DBSCAN yang mampu mendeteksi outlier atau Hierarchical Clustering, guna menilai konsistensi pola klaster yang terbentuk lintas metode.

4. Pemantauan lahan pertanian: Pemerintah daerah disarankan memanfaatkan klaster Kawasan Tanaman Pangan yang teridentifikasi sebagai basis pemantauan berkala terhadap potensi alih fungsi lahan sawah, khususnya bidang-bidang tanah yang berbatasan langsung dengan kawasan permukiman perkotaan.

DAFTAR PUSTAKA

- Akbar, F. (2024). Eksplorasi Teknologi Remote Sensing dalam Ilmu Ukur Tanah untuk Pengelolaan Lahan. *IlmuTeknik.org*, 3(3).
- Almais, A.T.W., et al (2023). Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters. *International Journal of Disaster Risk Reduction*, 11, 74590-74601.
- Utnasari, I. (2021). Analisis Clustering Dengan K-Means Untuk Pengelompokan Penjualan Produk Pada Hotel Newton. *Jurnal Comasie*, 4(4), 1-8.
- Fithriyah, M., Yaqin, M. A., Zaman, S. (2021). K-Means Clustering Untuk Segmentasi Produk Berdasarkan Analisis Recency, Recency, Recency (RFM) pada Data Transaksi Penjualan. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 3(2), 151-164.
- Ilham, A., Assaffat, L., Khikmah, L., & Suprapedi. (2023). k-Means Cluster-based Random Undersampling and Meta-Learning Approach for Village Development Status Classification. *JOIV : International Journal on Informatics Visualization*, 7(2), 542-547.
- Keputusan Menteri Agraria dan Tata Ruang / Badan Pertanahan Nasional Nomor 1589/SK-HK.02.01/XII/2021
- Peraturan Menteri Agraria dan Tata Ruang / Badan Pertanahan Nasional Nomor 13 Tahun 2021
- Peraturan Menteri Agraria dan Tata Ruang / Badan Pertanahan Nasional Nomor 14 Tahun 2021
- Peraturan Menteri Koordinator Bidang Perekonomian Nomor 18 Tahun 2020
- Peraturan Presiden Nomor 18 Tahun 2020
- Shie, L., & Maulana, A. F. (2024). Evaluasi Kebijakan Agraria Dalam Mendukung Keberlanjutan Lingkungan Dan Ekonomi. *Journal of Internasional Multidisciplinary Research*, 2(2), 20-25.
- Siregar, P. P., Solikhun, & Siregar, Z. A. (2022). Penerapan Metode K-Means Dalam Mengelompokkan Persebaran Lahan Kritis Di Indonesia Berdasarkan Provinsi. *RESOLUSI : Rekayasa Teknik Informatika dan Informasi*, 2(4), 145-151.
- Sugoto, H. A., Maarif, M. S., Mulyanto, B., & Harianto. (2024). Optimalisasi Peran Pemerintah Melalui Badan Bank Tanah sebagai Land Manager Negara dalam Pengelolaan Penyediaan Tanah bagi Kepentingan Pembangunan. *Jurnal Aplikasi Manajemen dan Bisnis*, 10(1).
- Suryo, M. S. (2017). Analisa Kebutuhan Luas Minimal pada Rumah Sederhana Tapak di Indonesia. *Jurnal Permukiman*, 12(2), 116-123.

- Syaputri, D., Noprita, P. H., & Romelah, S. (2021). Implementation of K-Means Algorithm for Economic Distribution Clustering Based on Demographics of Population. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 1-6.
- Trocolli, E. B., et al (2022). K-means Clustering Using Principal Component Analysis to Automate Label Organization in Multi-Attribute Seismic Facies Analysis. *Journal of Applied Geophysics*, 198(3). 104555. 1-17.