

Identifikasi Pola Konflik Lahan Perkebunan di Lingkungan PTPN Group Berbasis Data Hukum Menggunakan Hierarchical Clustering dengan Algoritma Agglomerative

Theresia Dwiati Wismarini¹, Sri Eniyanti², Endang Lestariningsih³, R. Soelistijadi⁴, Eka Ardhianto⁵

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Industri, Universitas Stikubank

⁵Program Studi Teknologi Informasi, Fakultas Teknologi Informasi dan Industri, Universitas Stikubank

¹thwismarini@edu.unisbank.ac.id, ²eniyati@edu.unisbank.ac.id, ³endanglestariningsih@edu.unisbank.ac.id,

⁴r.soelistijadi@edu.unisbank.ac.id, ⁵ekaardhianto@edu.unisbank.ac.id*

Abstract

It is very important to detect land conflicts early, because of their potential impact on various aspects such as health, ecosystems, agriculture and electricity networks. This study aims to identify patterns of land conflicts within PTPN Group plantations using legal data through Hierarchical Clustering with the Agglomerative Clustering algorithm. Early detection of land conflicts is crucial due to their impacts on health, ecosystems, agriculture, and infrastructure. The study utilizes legal data from the Supreme Court of Indonesia, geospatial data from OpenStreetMap, and socio-economic data from BPS and the World Bank to analyze and cluster conflict patterns. The analysis process includes data initialization, distance calculation, cluster merging, and dendrogram visualization. Results demonstrate that the algorithm effectively clusters land conflicts based on different population and GDP characteristics, helping to elucidate relationships between legal cases. This study contributes by identifying primary drivers of land conflicts, supporting data-driven land management. Public policy recommendations include establishing priority zones for conflict resolution, optimizing data-driven legal monitoring, and enhancing transparency in land governance to prevent conflict escalation in PTPN Group areas.

Keywords: plantation land conflicts, agglomerative clustering, legal data, geospatial data, socio-economic data

Abstrak

Konflik lahan perkebunan sangat penting untuk dideteksi secara dini, karena potensi dampaknya terhadap berbagai aspek seperti kesehatan, ekosistem, pertanian, dan jaringan listrik. Penelitian ini bertujuan mengidentifikasi pola konflik lahan perkebunan di lingkungan PTPN Group berbasis data hukum menggunakan teknik *Hierarchical Clustering* dengan algoritma *Agglomerative Clustering*. Deteksi dini konflik lahan penting karena dampaknya terhadap kesehatan, ekosistem, pertanian, dan infrastruktur. Studi ini mengolah data hukum dari Mahkamah Agung RI, data geospasial dari *OpenStreetMap*, dan data sosial-ekonomi dari BPS dan *World Bank* untuk menganalisis dan mengelompokkan pola konflik. Proses analisis meliputi inialisasi data, penghitungan jarak, penggabungan klaster, dan visualisasi dendrogram. Hasilnya menunjukkan bahwa algoritma ini efektif dalam mengelompokkan konflik lahan berdasarkan karakteristik populasi dan GDP yang berbeda, membantu memahami hubungan antar kasus hukum. Penelitian ini berkontribusi dengan mengidentifikasi faktor utama pemicu konflik lahan untuk mendukung manajemen lahan berbasis data. Rekomendasi kebijakan publik mencakup penetapan zona prioritas penyelesaian konflik, optimalisasi pengawasan hukum berbasis data, dan peningkatan transparansi dalam tata kelola lahan guna mencegah eskalasi konflik di wilayah PTPN Group.

Kata kunci: konflik lahan, agglomerative clustering, data hukum, data geospasial, data sosioekonomi

©This work is licensed under a Creative Commons Attribution -ShareAlike 4.0 International License

1. Pendahuluan

Deteksi dini konflik lahan perkebunan sangat penting karena potensi dampaknya terhadap berbagai aspek seperti kesehatan, ekosistem, pertanian, dan jaringan listrik. Mendeteksi konflik dengan segera memungkinkan penerapan langkah-langkah pengendalian yang efektif, mencegah eskalasi lebih lanjut dari masalah tersebut. Misalnya, dalam konteks konflik vegetasi dengan jaringan listrik [1], deteksi dini melalui model prediktif dapat membantu memprioritaskan area untuk inventarisasi vegetasi, mengurangi pemadaman listrik dan risiko bagi populasi. Selain itu, penggunaan teknologi penginderaan jauh seperti citra satelit dan data UAV dapat menyediakan pemantauan yang lebih terperinci

terhadap biomassa perkebunan, yang mendukung upaya mitigasi perubahan iklim [2]. Oleh karena itu, deteksi dini konflik di lahan perkebunan tidak hanya melindungi lingkungan dan keanekaragaman hayati tetapi juga memastikan pengelolaan sumber daya dan infrastruktur yang berkelanjutan, yang pada akhirnya menguntungkan ekosistem alami dan sistem sosial-ekonomi.

Deteksi dini konflik lahan perkebunan dapat secara signifikan membantu dalam pengambilan keputusan kebijakan dengan mencegah eskalasi dan mengatasi masalah mendasar secara cepat. Mendeteksi konflik lebih awal memungkinkan intervensi tepat waktu, mengurangi risiko perselisihan yang berkepanjangan dan dampak negatif potensial pada masyarakat dan

lingkungan [3]. Memanfaatkan teknologi penginderaan jauh, seperti pencitraan hiperspektral dan algoritma pembelajaran mesin, dapat meningkatkan deteksi konflik dengan memetakan perubahan penggunaan lahan dan mengidentifikasi area yang menjadi perhatian [4, 5]. Selain itu, mengukur sensitivitas sistem pengawasan deteksi dini dapat membantu mengevaluasi efektivitas berbagai pendekatan dalam mengidentifikasi konflik dan memastikan respons yang cepat [6]. Dengan mengintegrasikan metode-metode ini, pembuat kebijakan dapat secara proaktif menangani konflik lahan perkebunan, melindungi hak-hak pekerja, dan mempromosikan praktik pengelolaan lahan yang berkelanjutan, yang pada akhirnya mengarah pada keputusan kebijakan yang terinformasi dan efektif.

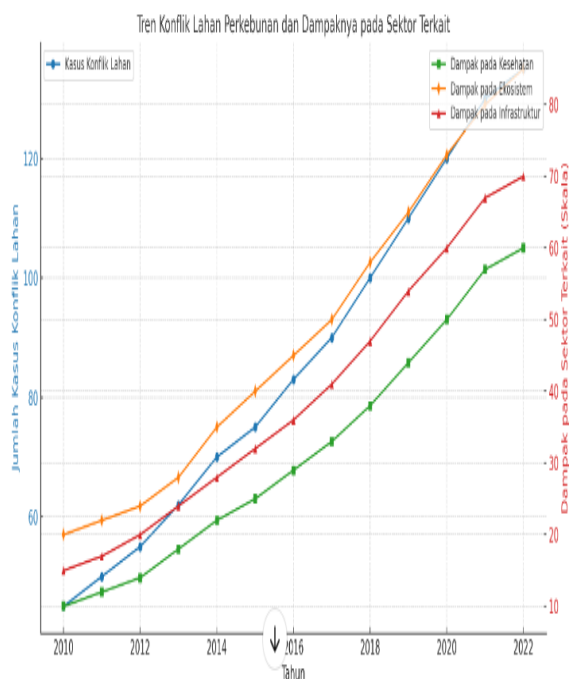
Kegagalan mendeteksi konflik lahan perkebunan sejak dini dapat menyebabkan berbagai dampak negatif, seperti perselisihan yang berkepanjangan antara komunitas dan perusahaan, mekanisme resolusi konflik yang bias dan lebih menguntungkan perusahaan, ketidakseimbangan antara ekosistem alam dan sistem sosial-ekonomi, serta meningkatnya konflik antara satwa liar dan pertanian. Deteksi konflik yang tertunda dapat mengakibatkan terbatasnya akses keadilan bagi komunitas pedesaan, sehingga konflik lahan terus berlanjut [7]. Selain itu, hal ini dapat menghambat pengelolaan sumber daya air dan penggunaan lahan yang efektif, mengganggu keseimbangan ekologi dan produktivitas [8]. Lebih jauh lagi, mengabaikan konflik antara satwa liar dan pertanian dapat menyebabkan kerusakan yang signifikan pada lahan yang dibudidayakan, mempersulit upaya konservasi dan pertimbangan ekonomi [9]. Deteksi dini konflik lahan perkebunan sangat penting untuk mencegah perselisihan yang berkepanjangan, menjaga keseimbangan ekologi, dan memastikan keberlangsungan hidup yang berkelanjutan antara komunitas, perusahaan, dan satwa liar.

Indikator awal untuk mendeteksi konflik di lahan perkebunan dapat mencakup variasi dalam sifat fisikokimia tanah, seperti kelembaban tanah, total nitrogen (TN), dan karbon organik tanah (SOC), serta perubahan dalam biomassa dan komposisi mikroba tanah, seperti rasio kelompok mikroba yang berbeda (Jamur terhadap Bakteri, Gram-positif terhadap Gram-negatif). Selain itu, pencitraan hiperspektral dari kendaraan udara tak berawak (UAV) yang dikombinasikan dengan algoritma pembelajaran mesin dapat membantu mengklasifikasikan kategori infeksi yang berbeda pada tanaman, seperti penyakit busuk pangkal batang (BSR) pada pohon kelapa sawit. Lebih jauh lagi, representativitas spasial dari pengukuran *Eddy Covariance* (EC) di perkebunan hutan dapat memberikan wawasan tentang potensi konflik dengan mengevaluasi fraksi jenis vegetasi target (FTVT) dan bias lokasi sensor, yang dapat menunjukkan dampak penggunaan lahan pada pertukaran karbon, air, dan energi dalam ekosistem [5, 10, 11].

Teknologi data *science* memainkan peran penting dalam deteksi dini konflik lahan perkebunan dengan memanfaatkan berbagai teknik dan sumber data. Misalnya, dalam konteks pemantauan perkebunan kelapa sawit, pencitraan hiperspektral dari kendaraan udara tak berawak (UAV) yang dikombinasikan dengan algoritma pembelajaran mesin seperti random forest dapat mengklasifikasikan kategori infeksi penyakit yang mempengaruhi pohon dengan akurat, memungkinkan deteksi dini [5]. Selain itu, integrasi data optik satelit, radar aperture sintetis (SAR), dan UAV dapat digunakan untuk memperkirakan biomassa di atas permukaan tanah (AGB) di perkebunan mangrove, membantu pemantauan perubahan dari waktu ke waktu dan menilai potensi penyerapan karbon [1]. Dengan membangun model hidrologi yang didistribusikan berdasarkan unit penggunaan lahan dan modul air tanah, data science dapat mensimulasikan proses produksi dan konsumsi air, memberikan wawasan tentang interaksi antara sumber daya air dan lahan untuk mengelola konflik secara efektif [8].

Data hukum dapat memainkan peran penting dalam memahami penyebab utama konflik lahan dengan memberikan wawasan tentang kerangka hukum dan lembaga yang terlibat dalam penyelesaian sengketa. Perangkat lunak analitik hukum dapat membantu dalam memvisualisasikan dan menganalisis hubungan dalam yurisprudensi, mengungkap struktur dasar sistem hukum [12]. Selain itu, supremasi hukum dan kebebasan berkumpul ditemukan memiliki hubungan positif dengan kepuasan terhadap kebebasan, terutama di negara-negara demokrasi dan berpenghasilan tinggi, yang menyoroti pentingnya kebebasan hukum dalam mendorong kepuasan terhadap pilihan individu [13]. Lebih jauh lagi, analisis dataset pembelajaran mesin hukum dapat mengungkap ketidaksepakatan dan penanganannya selama proses anotasi, menawarkan saran untuk meningkatkan pelaporan anotasi yang bertentangan dalam dataset hukum [14]. Dengan memanfaatkan dataset tentang peristiwa keadilan transisional personalia, peneliti dapat mengukur tingkat keparahan, urgensi, dan volatilitas pendekatan negara dalam menangani pelanggaran hak asasi manusia di masa lalu, memberikan wawasan berharga tentang dampak penanganan mantan elit otoriter terhadap stabilitas demokrasi [15].

Konflik lahan perkebunan merujuk pada perselisihan yang timbul dari konversi lahan untuk perkebunan industri, yang berdampak pada komunitas lokal dan ekosistem. Di Indonesia, konflik sering kali berasal dari ekspansi perkebunan kelapa sawit, yang mengakibatkan deforestasi dan ancaman ekologis [16].



Gambar 1. Grafik Tren Konflik Lahan

Gambar 1 menunjukkan tren konflik lahan perkebunan di Indonesia dari tahun 2010 hingga 2023, serta dampaknya terhadap sektor-sektor terkait, yaitu Tren Konflik Lahan yang terjadi peningkatan jumlah kasus konflik lahan secara signifikan selama periode tersebut, dari 45 kasus pada tahun 2010 menjadi 135 kasus pada 2023. Konflik lahan ini memberikan dampak pada beberapa sektor, yaitu: 1) Sektor Kesehatan: Dampak pada kesehatan menunjukkan peningkatan konsisten, menyoroti risiko kesehatan yang dihadapi masyarakat akibat konflik seperti stres dan gangguan akses layanan kesehatan. 2) Sektor Ekosistem: Kerusakan ekosistem meningkat tajam, menunjukkan deforestasi dan hilangnya biodiversitas sebagai dampak utama konflik lahan, dan 3) Sektor Infrastruktur: Konflik berdampak pula pada kerusakan infrastruktur, termasuk jalan dan fasilitas publik, yang menghambat aksesibilitas masyarakat.

Roundtable on Sustainable Palm Oil (RSPO) telah menetapkan mekanisme penyelesaian konflik untuk menangani keluhan antara komunitas pedesaan dan perusahaan kelapa sawit, tetapi cenderung lebih menguntungkan perusahaan, sehingga konflik terus berlanjut [7]. Selain itu, praktik perladangan berpindah di Indonesia menghadapi tantangan karena akuisisi lahan besar-besaran oleh perusahaan, memaksa komunitas untuk beradaptasi dengan produksi komersial dan menyebabkan konflik atas penggunaan dan ketersediaan lahan [17]. Dengan memungkinkan agroforestri di dalam perkebunan komersial, seperti agroforestri pinus-kopi di Indonesia, konflik dapat diminimalkan, memberikan peluang pendapatan bagi komunitas pedesaan dan mengurangi tekanan pada hutan alami [18].

Mengumpulkan dan menganalisis data hukum terkait konflik lahan menghadapi tantangan signifikan, termasuk kebutuhan pelatihan analitik data lanjutan di kalangan akademisi hukum [12]. Selain itu, data yang hilang akibat kondisi cuaca buruk dan kerusakan sensor selama pengumpulan data penginderaan jauh dapat menghambat studi perubahan penggunaan lahan atau tutupan lahan (LULC) yang komprehensif [19]. Lebih jauh lagi, interpretasi dokumen hukum seperti kontrak dan undang-undang dapat bervariasi antar individu, yang menyebabkan ketidaksepakatan yang sering kali tidak ditangani dengan memadai dalam kumpulan data pembelajaran mesin hukum, sehingga mempengaruhi keandalan proses anotasi [20]. Selain itu, sensitivitas topik penelitian, seperti isu-isu terkait konflik dalam konservasi, dapat mengakibatkan bias nonrespons dan kekhawatiran etis, yang mempengaruhi kualitas dan akurasi data yang dikumpulkan dari kuesioner dan wawancara dalam penelitian konservasi [21].

Hierarchical clustering adalah metode yang banyak digunakan dalam analisis data yang mengorganisir data ke dalam struktur seperti pohon berdasarkan kesamaan. *Agglomerative clustering*, jenis dari *hierarchical clustering*, dimulai dengan setiap titik data sebagai kluster terpisah dan menggabungkan kluster yang paling dekat secara iteratif hingga semua titik data termasuk dalam satu kluster. Proses ini menciptakan dendrogram yang menggambarkan hubungan antara titik data. Algoritma ini menghitung jarak antara kluster menggunakan berbagai metode seperti *Unweighted Pair-Group Method with Arithmetic Mean (UPGMA)* [22]. Selain itu, pendekatan baru seperti *Centroid Auto-Fused Hierarchical Fuzzy C-Means (CAF-HFCM)* mengotomatisasi pembentukan hierarki kluster tanpa mengandalkan indeks validitas kluster, menawarkan jumlah kluster yang optimal dan mengurangi sensitivitas terhadap inisialisasi [23]. Dalam ilmu biomedis, *Hierarchical Density Clustering (HAL-x)* secara efisien memprediksi beberapa populasi sel dalam data sel tunggal, mencapai akurasi tinggi dalam mengklasifikasikan status klinis berdasarkan profil sel [24].

Teknik *hierarchical clustering*, seperti algoritma *Agglomerative clustering*, dapat diterapkan juga untuk mengidentifikasi pola konflik lahan perkebunan berdasarkan data hukum dengan menganalisis hubungan dan kesamaan antara berbagai kasus atau sengketa hukum. Metode ini dapat membantu mengelompokkan kasus-kasus serupa untuk mengungkap tren umum atau faktor-faktor yang berkontribusi pada konflik, sehingga membantu dalam memahami lanskap hukum yang kompleks [22, 25, 26]. Dengan memanfaatkan *hierarchical clustering*, peneliti dapat menangani masalah ketidakunikan dalam *clustering* ketika berhadapan dengan jarak yang sama, memastikan representasi data yang lebih akurat dan berpotensi mengungkap pola tersembunyi dalam dokumen hukum terkait konflik lahan perkebunan [22]. Selain itu, kemampuan algoritma untuk menangani

skenario kompleks dan menganalisis struktur yang rumit dapat memberikan wawasan berharga tentang sifat dan dinamika sengketa lahan perkebunan, menawarkan pendekatan berbasis data untuk mempelajari dan menyelesaikan konflik di domain ini [25].

Algoritma *agglomerative clustering* telah menunjukkan fleksibilitasnya dalam berbagai bidang penelitian, termasuk pengelolaan hutan dan analisis data *multiview*. Misalnya, algoritma ini digunakan dalam penelitian pada pola distribusi perkebunan Larix Kaempferi, yang menganalisis dampak berbagai strategi pengelolaan terhadap distribusi pohon di area perkebunan [27]. Selain itu, teknik ini diterapkan dalam klasifikasi spesies tanaman berdasarkan karakteristik fenotip, mendukung penelitian biodiversitas dengan hasil *clustering* yang akurat berbasis sampel probabilistik [26]. Namun, penelitian terdahulu cenderung berfokus pada data yang lebih terstruktur dan homogen, sementara konflik lahan perkebunan di lingkungan hukum dan sosial-ekonomi memiliki karakteristik yang kompleks dan heterogen. Hingga kini, studi yang menerapkan teknik *clustering* pada data hukum terkait konflik lahan masih terbatas pada pendekatan deskriptif atau menggunakan teknik *clustering* seperti K-Means dan DBSCAN yang memiliki keterbatasan dalam menangani data yang bervariasi dan tidak berstruktur.

K-Means, misalnya, memerlukan penentuan jumlah kluster sejak awal dan kurang optimal untuk data hukum dan sosial-ekonomi yang heterogen, sementara DBSCAN efektif dalam mendeteksi kluster berdensitas tinggi tetapi terbatas dalam menangani distribusi kasus hukum yang tersebar luas secara geografis. Di sisi lain, *Agglomerative clustering* memiliki keunggulan dalam membentuk struktur hierarkis tanpa memerlukan asumsi awal mengenai jumlah kluster. Hal ini memungkinkan algoritma untuk menyesuaikan diri dengan variasi data, terutama data *multiview* yang menggabungkan data hukum, geospasial, dan sosial-ekonomi dalam satu analisis terintegrasi. Metode ini juga membantu mengatasi masalah ketidakunikan dalam *clustering (non-unique clustering)* yang sering muncul dalam data hukum, seperti diidentifikasi pada [20, 22].

Selain itu, penelitian ini memperkenalkan pendekatan yang lebih maju dengan memanfaatkan algoritma yang dapat menangani data *multiview* tanpa langkah pasca-pemrosesan tambahan, seperti yang dikembangkan dalam penelitian terbaru yang mengusulkan *Agglomerative Neural Networks (ANN)* untuk *clustering multiview* [25]. Dengan demikian, penelitian ini memperluas cakupan *agglomerative clustering* ke domain baru, khususnya dalam identifikasi pola konflik lahan berbasis data hukum, yang sebelumnya belum dicapai oleh studi lain. Pendekatan ini memungkinkan identifikasi faktor-faktor utama pemicu konflik serta pemetaan wilayah yang rentan konflik di lingkungan

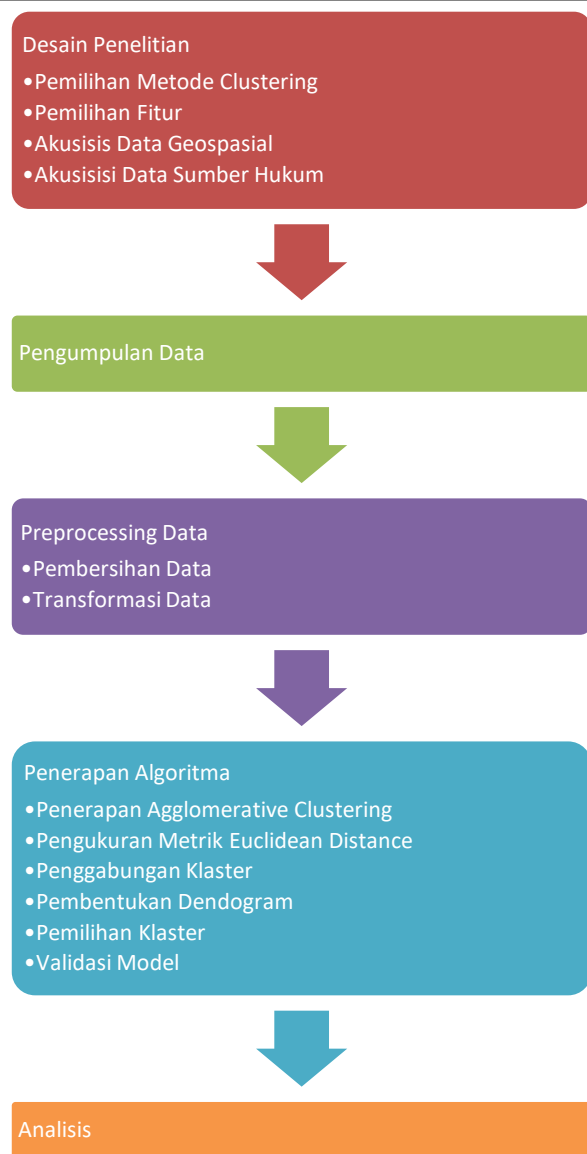
PTPN Group, yang belum disentuh oleh penelitian sebelumnya.

Dengan menggabungkan data hukum, geospasial, dan sosial-ekonomi, penelitian ini memberikan solusi konkret untuk mendukung kebijakan publik berbasis data melalui identifikasi wilayah prioritas untuk resolusi konflik, optimalisasi pengawasan hukum, dan peningkatan transparansi dalam tata kelola lahan.

Tujuan utama penelitian ini adalah untuk mengungkap tren umum dan faktor-faktor yang berkontribusi pada konflik lahan perkebunan dengan menganalisis hubungan dan kesamaan antara berbagai kasus atau sengketa hukum dan diharapkan dapat memberikan kontribusi signifikan dalam memahami dan mengelola konflik lahan perkebunan melalui pendekatan berbasis data hukum. Dengan menggunakan teknik *hierarchical clustering*, khususnya algoritma *agglomerative clustering*, penelitian ini akan mengidentifikasi pola dan tren umum dalam konflik lahan perkebunan. Hal ini akan membantu dalam mengungkap faktor-faktor yang berkontribusi pada terjadinya konflik, sehingga memberikan wawasan yang lebih mendalam tentang dinamika sengketa lahan. Selain itu, penelitian ini juga akan mengatasi beberapa celah dalam penelitian saat ini, seperti masalah ketidakunikan dalam *clustering hierarkis* akibat jarak yang sama. Dengan data science ini, penelitian ini akan meningkatkan akurasi dan efektivitas deteksi dini konflik lahan. Adapun secara awal, penelitian ini akan bermanfaat dalam kemajuan ilmu pengetahuan yaitu akan memperkaya literatur tentang penggunaan teknik *hierarchical clustering* dan algoritma *agglomerative clustering* dalam konteks analisis konflik lahan, serta mengatasi celah-celah yang ada dalam penelitian sebelumnya juga selain itu untuk peningkatan kapasitas akademis, dengan fokus pada analitik data lanjutan, penelitian ini akan meningkatkan kapasitas akademis dan pengetahuan di kalangan peneliti dan akademisi hukum terkait analisis data dan teknik *clustering*.

2. Metode Penelitian

Metodologi penelitian ini dirancang untuk mengidentifikasi pola konflik lahan perkebunan berbasis data hukum menggunakan teknik *hierarchical clustering*, khususnya algoritma *agglomerative clustering*. Bab ini akan menjelaskan secara rinci langkah-langkah yang ditempuh dalam penelitian, mulai dari desain penelitian, pengumpulan data, *preprocessing* data, hingga penerapan algoritma dan analisis hasil, yang terlihat pada gambar 2.



Gambar 2. Diagram Alir Metode Penelitian

2.1. Desain Penelitian.

Desain penelitian ini dirancang untuk secara sistematis mengidentifikasi pola konflik lahan perkebunan dengan menggunakan algoritma *agglomerative clustering*. Pada subbab ini, akan dijelaskan pendekatan yang digunakan untuk mengumpulkan, memproses, dan menganalisis data hukum, geospasial, dan sosial-ekonomi yang relevan. Desain penelitian ini mencakup pemilihan metode *clustering* yang tepat, penentuan fitur-fitur penting, serta langkah-langkah untuk memastikan validitas dan reliabilitas hasil yang diperoleh. Dengan pendekatan yang terstruktur dan metodis, penelitian ini bertujuan untuk menghasilkan wawasan yang mendalam dan aplikatif dalam memahami dinamika konflik lahan perkebunan.

Desain eksperimen dalam penelitian ini bertujuan untuk mengidentifikasi pola konflik lahan perkebunan dengan memanfaatkan data hukum, geospasial, dan sosial-ekonomi. Langkah pertama adalah mengumpulkan data dari berbagai sumber terpercaya

seperti Mahkamah Agung RI untuk data hukum, *OpenStreetMap* dan *USGS Earth Explorer* untuk data geospasial, serta Badan Pusat Statistik (BPS) dan *World Bank Data* untuk data sosial-ekonomi. Setelah data terkumpul, dilakukan proses pembersihan dan preprocessing untuk menghilangkan informasi yang tidak relevan atau duplikat. Data yang sudah bersih kemudian diubah menjadi fitur-fitur yang relevan untuk analisis lebih lanjut. Selanjutnya, algoritma *agglomerative clustering* diterapkan untuk mengelompokkan data berdasarkan kesamaan karakteristiknya, sehingga terbentuk pola-pola tren umum dan faktor-faktor yang berkontribusi pada konflik lahan. Hasil dari *clustering* ini divisualisasikan dalam bentuk dendrogram, yang membantu dalam menentukan jumlah kluster optimal dan memahami struktur hubungan antar kasus atau sengketa hukum. Dengan desain eksperimen ini, diharapkan diperoleh gambaran mendalam dinamika konflik lahan perkebunan dan faktor-faktor yang mempengaruhinya, sehingga dapat membantu dalam pengambilan keputusan kebijakan yang lebih efektif.

Metode pengumpulan data dalam penelitian ini melibatkan beberapa langkah yang sistematis dan terstruktur untuk memastikan kelengkapan dan keandalan data yang akan dianalisis. Untuk data hukum, sumber utama yang digunakan adalah Mahkamah Agung RI melalui portal <http://www.putusan3.mahkamahagung.go.id> dan situs Hukum online. Data dari sumber-sumber ini mencakup dokumen hukum seperti putusan pengadilan, peraturan perundang-undangan, dan kebijakan terkait lahan perkebunan.

Data geospasial diperoleh dari beberapa sumber terpercaya seperti *OpenStreetMap*, yang menyediakan peta lahan yang dapat diunduh secara gratis, dan *USGS Earth explorer*, yang menawarkan data citra satelit untuk analisis perubahan lahan. Selain itu, Badan Informasi Geospasial (BIG) menyediakan data geospasial Indonesia yang mencakup peta-peta tematik. Fokus pengumpulan data geospasial adalah pada wilayah Jawa dan Sumatera, yang memiliki lahan perkebunan yang signifikan.

Untuk data sosial-ekonomi, sumber utama adalah Badan Pusat Statistik (BPS) yang menyediakan data statistik Indonesia mencakup informasi demografi dan ekonomi, serta *World Bank Data* yang menawarkan data sosial-ekonomi global untuk perbandingan dan analisis. Data dari sumber-sumber ini mencakup informasi seperti aktivitas ekonomi, demografi, dan kondisi sosial-ekonomi daerah sekitar lahan perkebunan.

Setiap data yang dikumpulkan kemudian disimpan dalam format yang terstruktur dan siap untuk diproses lebih lanjut. Proses pengumpulan data ini bertujuan untuk menyediakan dataset yang komprehensif dan representatif, yang akan digunakan dalam analisis pola

konflik lahan perkebunan menggunakan algoritma *agglomerative clustering*.

2.2. Pengumpulan Data

Pengumpulan data merupakan langkah yang bertujuan untuk mengumpulkan informasi yang akurat dan relevan dari berbagai sumber. Data yang dikumpulkan meliputi data hukum, data geospasial, dan data sosial-ekonomi, yang akan digunakan untuk mengidentifikasi pola konflik lahan perkebunan. Dalam subbab ini, dijelaskan secara rinci metode dan proses pengumpulan data dari sumber-sumber terpercaya. Pendekatan yang sistematis dan terstruktur diharapkan dapat menghasilkan dataset yang komprehensif dan valid, mendukung analisis lebih lanjut dengan algoritma *agglomerative clustering*.

Terkait Sumber Data Hukum: Mahkamah Agung RI, Hukumonline. Data hukum diakses melalui direktori putusan pengadilan yang tersedia secara *online* di situs web Mahkamah Agung RI. Melakukannya adalah dengan *scraping* atau pengunduhan data dalam format yang sesuai, seperti CSV atau JSON, untuk diolah lebih lanjut. Selain itu, juga untuk mengakses artikel dan laporan dari Hukumonline, menggunakan API atau teknik *web scraping* untuk mendapatkan informasi terkait kasus-kasus hukum yang relevan dengan konflik lahan. Terkait Sumber Data Geospasial: *OpenStreetMap*, *USGS Earth Explorer*, Badan Informasi Geospasial (BIG). Untuk data geospasial ini memanfaatkan API dari *OpenStreetMap* untuk mengunduh data peta yang mencakup informasi penggunaan lahan dan infrastruktur. Dari *USGS Earth Explorer*, diunduh citra satelit dan data penginderaan jauh yang relevan dengan area penelitian, seperti Jawa dan Sumatera. Sementara itu, dari Badan Informasi Geospasial (BIG), mengakses data resmi geospasial Indonesia melalui portal data mereka, mengunduh peta topografi dan data penggunaan lahan untuk analisis lebih lanjut. Terkait Sumber Data Sosial-Ekonomi: Badan Pusat Statistik (BPS), *World Bank Data* memperoleh data sosial-ekonomi dengan mengakses portal data BPS, mengunduh data statistik resmi yang mencakup populasi, ekonomi, dan pertanian dalam format yang dapat diolah, seperti CSV. Dari *World Bank Data*, mengakses berbagai indikator sosial-ekonomi global menggunakan API atau teknik pengunduhan data yang tersedia, untuk membandingkan dan menganalisis data Indonesia dengan negara lain. Terkait Sumber Data Konflik Lahan: *Land Matrix*, Konsorsium Pembaruan Agraria (KPA), Untuk data konflik lahan, diakses *platform Land matrix* untuk mendapatkan data tentang akuisisi lahan skala besar, menggunakan API atau mengunduh dataset yang tersedia. Sedang dari Konsorsium Pembaruan Agraria (KPA), memperoleh laporan dan data tentang konflik agraria di Indonesia, menggunakan teknik *web scraping* atau mengunduh laporan dalam format PDF atau lainnya yang kemudian diubah menjadi format yang dapat diolah lebih lanjut.

2.3. Preprocessing Data

Bagian ini menjelaskan langkah-langkah preprocessing data yang mencakup pembersihan dan transformasi data untuk memastikan kualitas data yang digunakan dalam penelitian. Proses ini sangat penting agar analisis dapat menghasilkan temuan yang akurat dan andal. Langkah-langkah yang dilakukan meliputi:

1. Pembersihan Data:

Tahap pembersihan bertujuan untuk menghilangkan atau mengoreksi data yang tidak relevan, duplikat, atau berpotensi mengganggu analisis. Proses ini mencakup: 1) Pengecekan Konsistensi Data, dilakukan untuk memastikan semua data berada dalam format yang seragam. Misalnya, data hukum yang memiliki tanggal berbeda dari standar akan disesuaikan atau diperbaiki agar selaras dengan data lainnya. 2) Penghapusan Entri yang Tidak Lengkap, data yang memiliki informasi penting yang hilang atau kosong, seperti lokasi geografis yang tidak lengkap atau kasus hukum tanpa deskripsi yang cukup, dihapus agar tidak menyebabkan bias dalam analisis. Penghapusan dilakukan hanya jika data yang hilang tersebut bersifat esensial dan tidak dapat diperbaiki. 3) Koreksi Kesalahan Data. Jika ditemukan data yang mengandung kesalahan misalnya, entri yang salah ketik atau format angka yang tidak konsisten dilakukan koreksi untuk meningkatkan akurasi dataset.

2. Transformasi Data:

Setelah data dibersihkan, langkah berikutnya adalah mentransformasikan data hukum, geospasial, dan sosial-ekonomi menjadi format dan fitur yang siap dianalisis. Tahap transformasi ini meliputi: 1) Konversi Format Data. Data yang berasal dari sumber berbeda, seperti data hukum dalam format teks dan data geospasial dalam format shapefile, dikonversi ke format yang seragam (misalnya, CSV atau JSON) agar dapat diolah bersama. 2) Normalisasi Nilai. Nilai numerik seperti GDP dan populasi dinormalisasi untuk memastikan perbedaan skala antar variabel tidak mempengaruhi hasil analisis. 3) Ekstraksi Informasi Penting. Dari data hukum, informasi terkait kategori kasus, lokasi, dan tahun kasus diekstraksi untuk memudahkan proses clustering. Pada data geospasial, atribut seperti jenis penggunaan lahan dan koordinat geografis juga diekstraksi untuk mendukung analisis spasial.

Dengan menerapkan langkah-langkah pembersihan dan transformasi ini, *dataset* yang dihasilkan diharapkan siap untuk diolah dalam model clustering, memastikan bahwa data berkualitas tinggi dan relevan bagi tujuan penelitian ini.

2.4. Penerapan *Agglomerative Clustering*

Agglomerative clustering secara mendetail untuk mengidentifikasi pola konflik lahan perkebunan. Metode ini digunakan karena kemampuannya untuk mengelompokkan data ke dalam struktur hierarkis yang

dapat membantu mengungkap pola dan tren tersembunyi dalam data. Berikut adalah langkah-langkah yang terlibat dalam proses penerapan *agglomerative clustering*: 1) Inisialisasi. Pada tahap awal, setiap titik data dianggap sebagai kluster tersendiri. Hal ini memungkinkan algoritma untuk mulai dengan jumlah kluster yang maksimal, yang kemudian akan berkurang seiring penggabungan kluster. 2) Penghitungan Jarak. Pada tahap ini, penghitungan jarak antara setiap pasangan kluster dilakukan menggunakan *Euclidean Distance* sebagai metrik jarak yang dipilih. Metrik ini dipilih karena kemampuannya untuk mencerminkan perbedaan absolut antar variabel dalam ruang multidimensi, yang relevan dalam analisis konflik lahan yang melibatkan variasi geografis, sosial, dan ekonomi. Rumus *Euclidean Distance* antara dua titik data A dan B dalam ruang n-dimensi diberikan seperti terlihat dalam persamaan (1).

$$Euclidean\ distance = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} \quad (1)$$

Notasi A_i dan B_i adalah koordinat titik data A dan B pada dimensi ke- i dan n adalah jumlah dimensi. *Euclidean Distance* mengukur jarak lurus antara dua titik data, memberikan representasi yang akurat tentang perbedaan absolut di antara variabel-variabel penting, seperti GDP, populasi, dan koordinat geografis, yang relevan untuk analisis konflik lahan. *Euclidean Distance* sangat cocok dalam penelitian ini dikarenakan memudahkan dalam mengidentifikasi perbedaan spasial dan sosio-ekonomi antar wilayah konflik. Dengan menggunakan *Euclidean Distance*, analisis dapat secara efektif mengelompokkan wilayah-wilayah yang memiliki perbedaan signifikan, memungkinkan pengelompokan wilayah konflik lahan berdasarkan variabel absolut.

Metrik ini dipilih dibandingkan dengan *Manhattan Distance* atau *Cosine Similarity* karena karakteristik data penelitian ini, yang lebih membutuhkan representasi jarak absolut daripada jarak per segmen atau kesamaan arah. *Manhattan distance* lebih cocok untuk data dengan variasi besar dalam dimensi individual, sementara *Cosine Similarity* lebih fokus pada kesamaan arah dan kurang sesuai untuk data spasial dan sosio-ekonomi yang kompleks. Oleh karena itu, *Euclidean distance* memberikan dasar yang lebih kuat dalam menganalisis dan mengidentifikasi pola konflik lahan berdasarkan variasi geografis dan ekonomi yang absolut.

1. Penggabungan Kluster

Menggabungkan dua kluster yang memiliki jarak minimum. Proses ini berulang hingga semua titik data tergabung dalam satu kluster atau jumlah kluster yang diinginkan tercapai. Misalnya, jika kita menggunakan jarak *Euclidean*, kita akan menggabungkan dua kluster C_i dan C_j dengan jarak minimum.

2. Pembentukan Dendrogram

Membentuk dendrogram yang menggambarkan struktur kluster yang bersarang. Dendrogram ini membantu dalam menentukan jumlah kluster optimal. Setiap cabang dalam dendrogram menunjukkan penggabungan dua kluster pada titik waktu tertentu dalam proses *clustering*.

3. Pemilihan Jumlah Kluster

Memotong dendrogram pada level tertentu untuk menentukan jumlah kluster optimal. Level pemotongan dipilih berdasarkan kriteria seperti jarak antar kluster atau jumlah kluster yang diinginkan.

4. Validasi Model

Mengevaluasi hasil *clustering* menggunakan metrik validasi seperti *silhouette score* atau *Davies-Bouldin index* untuk memastikan representasi data yang lebih akurat dan mengungkap pola tersembunyi. *Silhouette score*, misalnya, dihitung menggunakan persamaan (2).

$$Silhouette\ Score = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

dimana $a(i)$ adalah rata-rata jarak antara titik data i dan semua titik data dalam kluster yang sama, sedangkan $b(i)$ adalah rata-rata jarak antara titik data i dan semua titik data dalam kluster terdekat yang berbeda.

5. Interpretasi Hasil

Menganalisis hasil *clustering* untuk mengidentifikasi pola konflik dan faktor-faktor yang berkontribusi pada konflik lahan perkebunan. Hasil ini kemudian digunakan untuk memberikan wawasan yang lebih dalam mengenai dinamika sengketa lahan, yang dapat menjadi dasar untuk rekomendasi kebijakan yang lebih efektif.

Dengan mengikuti langkah-langkah ini, kita dapat menerapkan *agglomerative clustering* untuk secara efektif mengidentifikasi dan menganalisis pola konflik lahan perkebunan berdasarkan data yang ada.

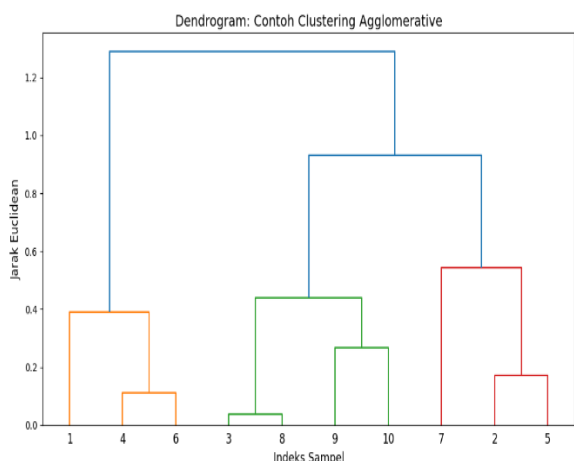
2.5. Analisis Data

Pada bagian ini, kita akan melakukan analisis data yang mencakup dua komponen utama: analisis hasil *clustering* untuk mengidentifikasi pola konflik dan interpretasi hasil untuk mengidentifikasi fitur-fitur penting yang berkontribusi pada konflik lahan. Dengan demikian, kita dapat memperoleh wawasan yang lebih mendalam tentang dinamika konflik lahan dan faktor-faktor yang mempengaruhinya yaitu: 1) Analisis Hasil *Clustering* untuk Mengidentifikasi Pola Konflik, dan 2) Interpretasi Hasil dan Identifikasi Fitur Penting.

Pada analisis hasil *clustering* untuk mengidentifikasi pola konflik. Langkah pertama dalam analisis data adalah mengevaluasi hasil dari proses *clustering* yang telah dilakukan. Dalam hal ini, kita akan melihat bagaimana data dikelompokkan berdasarkan kesamaan karakteristik. Dengan menggunakan dendrogram seperti terlihat dalam gambar 3, akan teridentifikasi kelompok-kelompok (kluster) yang terbentuk dan memahami bagaimana konflik lahan tersebar di berbagai wilayah. Pola-pola yang muncul dari hasil

clustering ini memberikan indikasi tentang area-area yang rentan terhadap konflik dan dapat membantu dalam menentukan prioritas penanganan.

Dalam interpretasi hasil dan identifikasi fitur penting. Setelah mengidentifikasi kluster-kluster yang ada, langkah selanjutnya adalah menginterpretasikan hasil ini untuk memahami karakteristik masing-masing kluster. Akan diperlihatkan juga secara rinci fitur-fitur yang menonjol di setiap klusternya. Dengan identifikasi fitur-fitur penting ini membantu dalam mengungkapkan faktor-faktor kunci yang berkontribusi pada konflik lahan dan dengan pemahaman ini, dapatlah dikembangkan strategi yang lebih tepat sasaran untuk mengelola dan meredam konflik lahan.



Gambar 3. Contoh Dalam Grafik Dendrogram

Adapun melalui gambar 3 tersebut akan dijelaskan bahwa hasil clustering akan terbentuk dengan 3 tahap yaitu: pertama adalah **Inisialisasi Kluster**, di mana pada awalnya, setiap data individu dianggap sebagai kluster tersendiri. Dalam dendrogram, ini direpresentasikan oleh ujung bawah setiap cabang (level paling bawah). Tahap Kedua dengan **Penggabungan Kluster**, yang mana akan mencari dua kluster yang paling dekat berdasarkan jarak Euclidean. Dua kluster yang memiliki jarak minimum digabungkan. Misal sampel 4 dan 6 memiliki jarak terkecil, sehingga digabung terlebih dahulu. Tahap ketiga **Hierarki Kluster**, kluster yang terbentuk digabungkan kembali secara bertahap dengan kluster lainnya. Proses ini diulang hingga semua kluster tergabung menjadi satu hierarki besar, terlihat pada puncak dendrogram. Tahap terakhir adalah **Penentuan Jumlah Kluster**, di mana untuk menentukan jumlah kluster optimal, dendrogram dapat "dipotong" secara horizontal pada level tertentu. Misalnya: Jika dipotong pada jarak Euclidean 0.6, dendrogram menghasilkan **3 kluster utama**. Sedangkan data dikelompokkan menjadi **Kluster Oranye, yang mengelompokkan sampel 1, 4, dan 6** karena jarak antar data dalam kluster ini sangat kecil (berdasarkan jarak Euclidean). **Kluster Hijau**: Terdiri dari sampel 3, 8, 9, dan 10. Kluster ini terbentuk dari data dengan jarak menengah yang lebih besar

dibandingkan kluster oranye dan **Kluster Merah**: mengelompokkan sampel 2, 5, dan 7. Kluster ini memiliki jarak internal yang lebih kecil dibandingkan dengan hijau.

Dalam hal ini Dendrogram menunjukkan bahwa data dapat dipecah menjadi beberapa kelompok berdasarkan tingkat kemiripan (jarak *Euclidean*). Semakin rendah garis horizontal di dendrogram, semakin dekat jarak antar data di kluster tersebut. Juga potensi jumlah kluster dapat dipilih sesuai dengan kebutuhan analisis, misalnya, dengan memotong dendrogram pada ketinggian tertentu untuk fokus pada pola-pola tertentu dalam data.

3. Hasil dan Pembahasan

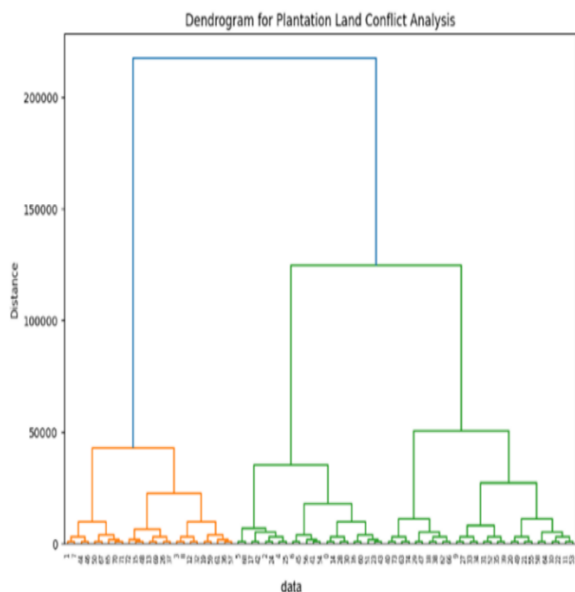
3.1. Hasil *Algoritma Agglomerative*

Algoritma Agglomerative diimplementasikan dalam program berbahasa python yang memiliki beberapa fungsi yang berurutan untuk mengumpulkan, membersihkan, dan menggabungkan data dari berbagai sumber sebelum menerapkan algoritma agglomerative clustering. Berikut adalah penjelasan langkah-langkah detail dari implementasi.

Pertama adalah pembuatan *dataset*, yang terdiri dari `create_hukum_data`, untuk dataset data hukum dari Mahkamah Agung, `create_sosial_ekonomi_data`, untuk dataset data sosial-ekonomi dari BPS, `create_geospasial_data`, untuk membuat dataset data geospasial menggunakan geometri poligon. Selanjutnya, pengolahan data, yang meliputi pengolahan `read_csv_with_auto_delimiter`: Membaca file CSV dengan deteksi delimiter otomatis, `get_mahkamah_agung_data`: Membaca dan membersihkan data hukum, `get_openstreetmap_data`: Mengunduh dan memproses data geospasial dari OpenStreetMap, `get_bps_data`: Mengunduh dan memproses data sosial-ekonomi dari BPS, `merge_data`: Menggabungkan data dari berbagai sumber menjadi satu dataset gabungan. Dan terakhir adalah, proses *agglomerative clustering* yang akan menerapkan langkah-langkah *agglomerative clustering* untuk mengidentifikasi pola konflik lahan Perkebunan seperti berikut pertama Inisialisasi, di mana setiap titik data dianggap sebagai kluster tersendiri. Kedua Penghitungan Jarak, yang akan Menghitung jarak antara setiap pasangan kluster menggunakan metrik *Euclidean*. Ketiga Penggabungan Kluster, yang menggabungkan dua kluster yang memiliki jarak minimum. Keempat Pembentukan Dendrogram, yang membentuk dendrogram untuk menggambarkan struktur kluster yang bersarang. Kelima Pemilihan Jumlah Kluster, yang memotong dendrogram pada level tertentu untuk menentukan jumlah kluster optimal. Keenam Validasi Model, yang mengevaluasi hasil clustering menggunakan metrik validasi seperti *silhouette score* dan *Davies-Bouldin index*. Ketujuh Interpretasi Hasil, yang menganalisis hasil *clustering* untuk mengidentifikasi pola konflik.

3.2. Analisis Grafik Dendrogram

Grafik dendrogram yang dihasilkan adalah representasi hierarkis dari hasil clustering menggunakan metode agglomerative clustering. Adapun grafik dendrogram ini akan diperlihatkan melalui gambar 4.



Gambar 4. Grafik Dendrogram Analisis Konflik Lahan Perkebunan.

Berikut beberapa poin utama dari apa yang diperlihatkan dalam grafik dendrogram ini. Pertama adalah Hierarki Kluster, yang mana Grafik Dendrogram menunjukkan bagaimana setiap titik data atau kluster digabungkan berdasarkan jarak yang dihitung menggunakan metode *linkage* (Ward's method). Pada bagian paling bawah dendrogram, setiap data adalah kluster tersendiri. Saat kita naik ke atas dendrogram, kluster-kluster ini digabungkan menjadi kluster yang lebih besar. Kedua jarak antar kluster, di mana sumbu vertikal (jarak) menunjukkan jarak atau *dissimilarity* antara kluster yang digabungkan dan garis biru panjang di bagian atas grafik dendrogram menunjukkan jarak besar antara dua kluster utama yang digabungkan. Ketiga jumlah kluster optimal, yang mana pemotongan grafik dendrogram pada level tertentu akan menentukan jumlah kluster optimal. Misalnya jika kita memotong dendrogram pada jarak 100000, kita akan mendapatkan tiga kluster utama (dua hijau dan satu oranye). Keempat Distribusi Kluster, yang mana grafik dendrogram ini menunjukkan bahwa terdapat beberapa sub-kluster kecil yang kemudian digabungkan menjadi kluster yang lebih besar. Distribusi kluster dapat membantu dalam memahami bagaimana data kelompok dan hierarki yang ada di antara data tersebut.

Berdasarkan dataset yang digunakan, berikut detail informasi terkait dataset yang digunakan beserta contoh valuedatanya yang membantu dalam analisis konflik lahan Perkebunan. Yaitu yang pertama terkait Data Hukum (*mahkamah_agung_data.csv*). Berisi informasi tentang kasus hukum terkait lahan dengan kolom seperti ID, *Case_Title*, *Date*, dan *Summary*. ID

untuk Identifikasi unik untuk setiap kasus, *Case_Title* merupakan Judul kasus, sedang *Date* adalah Tanggal kasus dan *Summary* merupakan ringkasan kasus. Contoh datanya misalnya, "Case A" dengan tanggal "2023-01-01" memiliki ringkasan "*Summary of Case A*", terlihat dalam tabel 1.

Tabel 1. Tabel Data Hukum

ID	Case Title	Date	Summary
1	Case A	2023-01-01	Summary of Case A

Kedua terkait Data Geospasial (*openstreetmap_data.shp*), Berisi informasi tentang penggunaan lahan dengan atribut seperti *osm_id*, *landuse*, *name*, dan *geometry*. Adapun *osm_id* mewakili Identifikasi unik untuk setiap lahan. *Landuse* merupakan Jenis penggunaan lahan (*farmland*, *orchard*, *plantation*). Kemudian *name* adalah Nama lahan dan *geometry* yaitu Geometri lahan (koordinat poligon). Contoh *value* datanya: "*Field A*" dengan penggunaan lahan "*farmland*" memiliki geometri berupa *polygon*, seperti terlihat dalam tabel2.

Tabel 2. Data Geospasial.

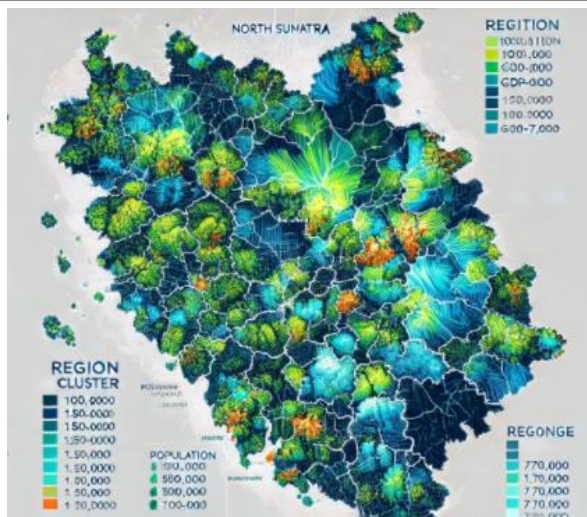
OSM_ID	Landuse	Name	Geometry
101	Farmland	Field A	Polygon A

Ketiga yaitu Data Sosial-Ekonomi (*bps_data_cleaned.csv*). Berisi data terkait populasi dan GDP untuk masing-masing wilayah yang memiliki atribut ID, *Region*, *Population*, dan GDP. ID adalah Identifikasi unik untuk setiap wilayah. Kemudian *Region* yaitu Nama wilayah, selanjutnya *population* adalah Populasi wilayah, GDP, singkatan dari Produk Domestik Bruto wilayah. Dengan contoh data adalah "*Region A*" memiliki populasi "100000" dan GDP "5000", seperti terlihat dalam tabel3.

Tabel 3. Sosial-Ekonomi.

ID	Region	Population	GDP
1	Region A	100000	5000

Kemudian terkait interpretasi hasil *Clustering* yang diperlihatkan, terdapat hal-hal berikut ini yaitu: pertama adalah untuk Identifikasi Pola Konflik, dari kluster yang terbentuk dapat menunjukkan wilayah atau kasus hukum yang memiliki kesamaan dalam hal populasi dan GDP, yang akan mengindikasikan adanya pola dalam konflik lahan berdasarkan faktor ekonomi dan demografi. Kedua penunjukan adanya Pengelompokan Berdasarkan Geospasial, di mana melalui Kluster ini juga dapat membantu mengidentifikasi wilayah geografis yang cenderung memiliki konflik serupa, berdasarkan penggunaan lahan (*farmland*, *orchard*, *plantation*). Sehingga implikasi untuk pengambilan keputusannya adalah dengan mengetahui kluster mana yang memiliki konflik serupa, pembuat kebijakan dapat merumuskan strategi yang lebih efektif untuk menyelesaikan konflik lahan dan mengelola sumber daya dengan lebih baik.



Gambar 5. Peta Kluster Konflik Lahan

Kluster yang dihasilkan dari grafik dendrogram menunjukkan pengelompokan berdasarkan fitur sosial-ekonomi, yaitu *Population* dan *GDP*. Kita melihat dua kluster utama yang terbentuk (hijau dan biru), walaupun terlihat juga kluster orange. Berikut merupakan rincian lebih spesifik terkait kluster-kluster tersebut. Yaitu: Pertama Kluster Hijau, adalah yang mempunyai karakteristik dengan wilayah populasi dan *GDP* yang lebih rendah. Data yang termasuk di dalamnya mempunyai kisaran data populasi: Sekitar 100.000 hingga 110.000 sedangkan untuk data *GDP*, sekitar 5.000 hingga 6.000. Contoh valuedata tersebut berisi diantaranya adalah ID: 1, Region: Region A, Population: 100000, *GDP*: 5000, ID: 2, Region: Region B, Population: 101000, *GDP*: 5100, ID: 10, Region: Region J, Population: 109000, *GDP*: 5900. Kedua adalah Kluster Biru, yang mana mempunyai karakteristik dengan Wilayah dengan populasi dan *GDP* yang lebih tinggi. Data yang termasuk di dalamnya adalah dengan kisaran data populasi sekitar 150.000 hingga 160.000 dan data *GDP*: Sekitar 10.000 hingga 11.000. Adapun contoh datanya adalah terdapat ID: 70, Region: Region Q, Population: 150000, *GDP*: 10000, ada ID: 71, Region: Region R, Population: 151000, *GDP*: 10100 dan ada ID: 75, Region: Region U, Population: 155000, *GDP*: 10500. Ketiga adalah Kluster Oranye dengan Karakteristik Wilayah yang mempunyai populasi dan *GDP* yang sedang (tidak terlalu rendah atau terlalu tinggi). Data yang termasuk di dalamnya data populasi berkisar sekitar 120.000 hingga 130.000 juga data *GDP* sekitar 7.000 hingga 8.000. Adapun contoh datanya adalah ID: 30, Region: Region X, Population: 120000, *GDP*: 7000, ID: 31, Region: Region Y, Population: 121000, *GDP*: 7100 dan ID: 40, Region: Region Z, Population: 130000, *GDP*: 8000. Hasil kluster ini diperlihatkan dalam gambar 5. yang menampilkan hubungan geospasial dengan penggunaan lahan (farmland, orchard, plantation) untuk membantu analisis konflik lahan dan pengambilan keputusan berbasis kluster.

Grafik dendrogram yang dihasilkan merupakan representasi hierarkis dari hasil clustering menggunakan metode agglomerative clustering, yang memungkinkan visualisasi pola konflik dalam data konflik lahan. Setiap cabang pada dendrogram menggambarkan proses penggabungan kluster secara bertahap berdasarkan tingkat kemiripan atau jarak antar data. Pada bagian paling bawah dendrogram, setiap titik data atau kasus konflik lahan direpresentasikan sebagai kluster tersendiri. Seiring dengan naiknya cabang dendrogram, titik-titik data ini mulai digabungkan dengan kluster lain yang paling mirip atau paling dekat jaraknya hingga membentuk kluster yang lebih besar.

Hierarki Kluster: Cabang-cabang di dendrogram mencerminkan struktur hierarkis dari hubungan antar data. Setiap kali dua cabang bertemu, hal tersebut menunjukkan adanya penggabungan dua kluster berdasarkan jarak yang dihitung. Semakin tinggi posisi cabang yang menghubungkan dua titik atau kluster, semakin besar jarak atau perbedaan antara kluster-kluster tersebut, menunjukkan bahwa penggabungan terjadi pada tingkat kesamaan yang lebih rendah. Pada konteks konflik lahan, cabang-cabang ini bisa mengindikasikan tingkat kemiripan pola konflik antara berbagai wilayah berdasarkan atribut sosial-ekonomi atau hukum yang dianalisis.

Jarak Antar Kluster: Sumbu vertikal (*distance*) pada dendrogram menunjukkan nilai jarak atau *dissimilarity* antara kluster-kluster yang digabungkan. Garis biru panjang di bagian atas dendrogram menunjukkan bahwa terdapat dua kluster utama dengan jarak yang besar di antara mereka, mengindikasikan adanya perbedaan signifikan dalam karakteristik wilayah konflik lahan. Hal ini berguna untuk mengidentifikasi kelompok wilayah dengan pola konflik yang berbeda secara signifikan satu sama lain.

Jumlah Kluster Optimal: Dengan memotong dendrogram pada jarak tertentu, kita dapat menentukan jumlah kluster optimal untuk analisis. Misalnya, dengan memotong pada jarak 100.000, kita bisa mendapatkan tiga kluster utama (dua hijau dan satu oranye). Dalam konteks konflik lahan, pemotongan ini dapat mengidentifikasi kelompok wilayah yang memiliki karakteristik konflik serupa, memudahkan dalam memahami distribusi konflik dan tingkat intensitas yang berbeda antar wilayah.

Distribusi Kluster: Dendrogram ini juga menunjukkan adanya sub-kluster kecil yang kemudian digabungkan menjadi kluster yang lebih besar. Distribusi kluster ini dapat mengindikasikan adanya kelompok wilayah dengan pola konflik yang lebih seragam di dalam satu kluster besar, tetapi tetap memiliki perbedaan di dalam sub-kluster. Hal ini membantu memahami hubungan antara berbagai wilayah dengan pola konflik lahan yang berbeda, seperti daerah yang mungkin lebih rentan terhadap konflik ekonomi versus daerah yang

memiliki permasalahan hukum terkait penggunaan lahan.

Secara keseluruhan, cabang-cabang pada dendrogram ini tidak hanya menunjukkan penggabungan kluster berdasarkan jarak tetapi juga membantu dalam memahami variasi pola konflik lahan perkebunan, di mana kluster besar dan kecil dapat dianalisis untuk menemukan karakteristik umum dan unik antar wilayah.

Adapun untuk Interpretasi setiap Klasternya dapatlah diperinci sebagai berikut : Untuk Kluster Hijau yang mana memiliki populasi dan GDP yang lebih rendah dibandingkan dengan kluster lainnya tersebut disimpulkan sebagai wilayah dengan sumber daya ekonomi yang terbatas dan populasi yang relatif kecil. Sedangkan Kluster Biru yang dimaknai memiliki populasi dan GDP yang lebih tinggi dibandingkan dengan kluster lainnya disimpulkan sebagai wilayah dengan sumber daya ekonomi yang lebih besar dan populasi yang relatif besar. Kemudian untuk Kluster Oranye yang dimaknai memiliki populasi dan GDP yang berada di tengah-tengah, tidak terlalu rendah dan tidak terlalu tinggi dari lainnya disimpulkan sebagai wilayah dengan sumber daya ekonomi dan populasi yang sedang. Sehingga di sini secara analisis wilayah-wilayah dalam kluster hijau mungkin memiliki konflik lahan yang berbeda karakteristiknya dibandingkan dengan kluster lainnya, dikarenakan keterbatasan sumber daya ekonomi yang mempengaruhi dinamika konflik. Kemudian wilayah-wilayah dalam kluster biru kemungkinan memiliki dinamika konflik yang lebih kompleks dan kemungkinan juga memiliki kapasitas yang lebih besar untuk manajemen konflik karena sumber daya ekonomi yang lebih besar. Sedangkan wilayah-wilayah dalam kluster oranye mungkin menunjukkan pola konflik yang berbeda lagi, yang mencerminkan karakteristik populasi dan GDP yang berada di tengah-tengah. Selanjutnya dengan penggambaran wilayah-wilayah yang masing-masing mencerminkan karakteristik populasi dan GDP baik Tinggi, sedang dan rendah ini juga akan memungkinkan penggolongan alternatif Tindakan yang akan dilakukan terhadap kluster-kluster wilayah ini, yaitu seperti berikut ini. Untuk Kluster Hijau yang menggambarkan Wilayah dengan Populasi dan GDP yang Lebih Rendah, secara analisisnya membantu mengidentifikasi area yang mungkin memerlukan perhatian lebih dalam pengelolaan sumber daya dan mitigasi konflik lahan. Kemudian untuk Kluster Biru yang menggambarkan Wilayah dengan Populasi dan GDP yang Lebih Tinggi, dapat digunakan untuk mengidentifikasi wilayah dalam kluster ini dalam pemberian wawasan penting untuk pengambilan keputusan yang lebih informatif dalam pengelolaan lahan perkebunan. Sedangkan Kluster Oranye yang lebih mempunyai wilayah-Wilayah dengan pola konflik yang berbeda dikarenakan punya karakteristik populasi dan GDP yang berada di tengah-tengah. Di sini terdapat pola konflik dalam kluster ini mungkin

berbeda dari kluster hijau dan biru, yang mana mencerminkan dinamika yang unik dalam konteks populasi dan ekonomi. Maka analisis lebih lanjut diperlukan untuk memahami bagaimana karakteristik ini mempengaruhi konflik lahan perkebunan dan pengelolaan sumber daya.

Dengan menganalisis dendrogram ini, kita dapat mengidentifikasi pola umum dan spesifik di berbagai kluster yang terbentuk berdasarkan karakteristik sosial-ekonomi dan demografi setiap wilayah. Hasil clustering yang diperlihatkan pada dendrogram memungkinkan kita untuk mengenali kelompok wilayah dengan karakteristik tertentu misalnya, kluster hijau yang cenderung memiliki populasi dan GDP yang lebih rendah. Wilayah-wilayah dalam kluster ini menunjukkan keterbatasan sumber daya ekonomi dan mungkin lebih rentan terhadap konflik yang dipicu oleh tekanan ekonomi atau kesenjangan dalam akses ke sumber daya. Pemahaman ini dapat digunakan oleh pembuat kebijakan untuk mengarahkan perhatian dan alokasi sumber daya yang lebih besar pada kluster seperti ini guna mengurangi potensi konflik.

Sebaliknya, wilayah yang berada dalam kluster biru, dengan populasi dan GDP yang lebih tinggi, menunjukkan kemampuan ekonomi yang lebih baik dan mungkin memiliki kapasitas untuk menangani konflik dengan lebih efisien. Namun, kompleksitas konflik di wilayah ini mungkin lebih tinggi, mengingat akses sumber daya yang lebih besar dapat memicu konflik baru terkait distribusi atau penggunaan lahan. Dengan mengetahui potensi konflik ini, pembuat kebijakan dapat merancang strategi yang fokus pada pemantauan regulasi dan penyusunan kebijakan tata kelola lahan yang lebih ketat di wilayah-wilayah ini untuk mencegah eskalasi konflik.

Wilayah-wilayah dalam kluster oranye, yang memiliki populasi dan GDP menengah, menunjukkan pola konflik yang berbeda yang memerlukan pendekatan manajemen yang fleksibel. Wilayah dalam kluster ini dapat mengalami tekanan konflik baik dari faktor ekonomi maupun hukum yang sedang berkembang. Dengan demikian, pembuat kebijakan dapat menerapkan strategi proaktif untuk wilayah ini, seperti kebijakan mitigasi konflik dan edukasi masyarakat untuk meningkatkan keterampilan manajemen lahan yang berkelanjutan.

3.3. Prediksi Konflik Masa Depan

Data dari hasil clustering ini juga dapat digunakan untuk memprediksi potensi konflik di masa depan. Berdasarkan pola yang ditemukan dalam setiap kluster, kita dapat membuat model yang dapat mengidentifikasi indikator awal konflik berdasarkan karakteristik yang ada di setiap kluster. Misalnya, jika wilayah-wilayah yang tergabung dalam kluster hijau mulai menunjukkan peningkatan kepadatan populasi atau penurunan GDP, wilayah tersebut dapat diprioritaskan untuk intervensi lebih awal karena potensi peningkatan risiko konflik. Dengan demikian, dendrogram ini berfungsi sebagai

alat pemetaan risiko yang dapat digunakan untuk memantau perkembangan wilayah-wilayah rentan dan mengidentifikasi kondisi yang dapat memicu konflik.

3.4. Upaya Pencegahan Konflik

Hasil analisis ini juga dapat menjadi dasar bagi pembuatan kebijakan pencegahan konflik yang lebih efektif dan terarah. Dengan mengetahui wilayah-wilayah yang memiliki risiko konflik tinggi berdasarkan karakteristik klaster, pemerintah dan pihak berwenang dapat menerapkan kebijakan yang bersifat preventif, seperti: 1) Pengawasan dan Alokasi Sumber Daya. Menargetkan wilayah dengan risiko konflik tinggi untuk program pengembangan ekonomi dan peningkatan akses ke sumber daya, terutama di wilayah dengan populasi dan GDP rendah. 2) Penegakan Hukum yang Tepat Sasaran. Mengimplementasikan kebijakan hukum yang lebih ketat di wilayah dengan GDP tinggi, di mana potensi konflik seringkali berhubungan dengan distribusi dan akses terhadap sumber daya yang lebih kompleks. Dan 3) Edukasi dan Program Pelatihan Masyarakat: Untuk wilayah dengan karakteristik sosial-ekonomi menengah, pembuat kebijakan dapat mendorong program edukasi terkait manajemen lahan berkelanjutan dan penguatan kapasitas lokal untuk mengelola konflik yang muncul. Secara keseluruhan, analisis dendrogram ini bukan hanya menggambarkan kondisi konflik lahan saat ini, tetapi juga memberikan wawasan prediktif yang dapat membantu mengidentifikasi potensi risiko konflik di masa depan. Dengan demikian, hasil ini dapat menjadi alat yang sangat berguna untuk merencanakan kebijakan dan program yang lebih responsif dan terukur dalam upaya mencegah konflik lahan yang lebih luas.

4. Kesimpulan

Penelitian ini bertujuan untuk mengidentifikasi pola konflik lahan perkebunan berbasis data hukum dengan menggunakan teknik *hierarchical clustering*, khususnya algoritma *agglomerative clustering*. Hasil penelitian menunjukkan bahwa algoritma ini mampu mengelompokkan data hukum, geospasial, dan sosial-ekonomi menjadi klaster-klaster yang menggambarkan karakteristik populasi dan GDP yang berbeda. Dengan mengidentifikasi pola-pola dalam konflik lahan, penelitian ini memberikan wawasan berharga tentang dinamika konflik yang terjadi di wilayah perkebunan.

Adapun kelebihan yang dapat ditemukan dalam penelitian ini adalah dari segi Pendekatan Komprehensif, penelitian ini mengintegrasikan berbagai jenis data, yaitu data hukum, geospasial, dan sosial-ekonomi, yang memberikan gambaran menyeluruh tentang konflik lahan. Dari segi Metodologi dapat dikatakan sistematis, ini dikarenakan metodologi yang digunakan dalam penelitian ini, termasuk *preprocessing* data dan penerapan algoritma *Agglomerative clustering*, dilakukan dengan langkah-langkah yang jelas dan terstruktur. Kemudian dari segi

Analisis Klaster adalah mendalam, yang mana ini karena penelitian ini tidak hanya mengidentifikasi klaster, tetapi juga memberikan interpretasi rinci tentang karakteristik masing-masing klaster, membantu dalam memahami faktor-faktor yang berkontribusi pada konflik lahan.

Namun kekurangannya pun di penelitian ini masih ada. Yaitu terkait tentang kualitas hasil *clustering* sangat bergantung pada kualitas data yang digunakan. Data yang tidak lengkap atau tidak akurat dapat mempengaruhi hasil analisis. Selanjutnya jumlah data yang Terbatas, yang mana penelitian ini hanya menggunakan 75 data sampel yang mungkin belum mencakup seluruh variasi konflik lahan yang ada. Juga terkait keterbatasan dalam validasi model. Yang mana validasi model *clustering* pada penelitian ini hanya menggunakan dua metrik yaitu *silhouette score* dan *Davies-Bouldin index*, yang mungkin belum cukup untuk mengevaluasi keseluruhan kinerja model.

Untuk penelitian mendatang, disarankan untuk:

1. Memperluas cakupan metrik validasi: Meningkatkan akurasi evaluasi hasil *clustering* dengan menambahkan metrik validasi lain seperti *Adjusted Rand Index (ARI)* dan *Calinski-Harabasz Index*. *ARI* dapat memberikan wawasan lebih mendalam mengenai kesesuaian hasil *clustering* terhadap label kategori yang mungkin sudah ada, sedangkan *Calinski-Harabasz Index* dapat menilai kepadatan dan pemisahan antar klaster, memberikan gambaran kualitas klaster secara lebih menyeluruh.
2. Mengeksplorasi metode *clustering* yang Lebih Kompleks: Menggunakan metode *clustering* yang lebih fleksibel dan kompleks, seperti *Gaussian Mixture Models (GMM)* dan *Spectral Clustering*, untuk menangkap pola yang lebih rumit dalam data konflik lahan. *GMM*, yang memungkinkan bentuk distribusi yang lebih variatif, dapat berguna dalam analisis data heterogen seperti ini. Sementara itu, *Spectral Clustering* efektif untuk data dengan struktur non-linear, yang sering muncul pada data spasial dan konflik lahan.
3. Memperluas Sumber dan Kualitas Data: Menambah cakupan data dari sumber yang lebih luas dan detail, seperti data satelit, citra drone, dan data lapangan yang lebih mutakhir. Menggunakan data dari lembaga seperti LAPAN, Badan Informasi Geospasial (BIG), dan sumber internasional seperti *Google Earth Engine* dapat memperkaya konteks spasial dalam analisis. Selain itu, data sosial-ekonomi tambahan dari Bank Dunia atau UNDP dapat membantu memberikan konteks yang lebih dalam terhadap faktor-faktor eksternal yang mungkin memengaruhi dinamika konflik lahan.
4. Melakukan Analisis Lanjutan pada Klaster yang Menengah (Oranye): Karena klaster ini mencerminkan wilayah dengan populasi dan GDP yang tidak terlalu rendah atau tinggi, dinamika konflik di klaster ini

mungkin berbeda dari klaster lainnya. Analisis lanjutan, termasuk eksplorasi perbedaan spasial dan temporal, diperlukan untuk memahami faktor-faktor unik yang memengaruhi konflik di wilayah-wilayah ini.

5. Integrasi dengan Analisis Spasial dan Temporal: Menggabungkan pendekatan spasial dan temporal dapat memberikan wawasan lebih komprehensif tentang evolusi konflik lahan dari waktu ke waktu. Pendekatan ini memungkinkan pemetaan pola konflik secara dinamis, yang dapat membantu dalam prediksi konflik di masa depan dan memungkinkan pembuatan kebijakan yang lebih responsif.

Daftar Rujukan

- [1] Z. Wang, L. Yi, W. Xu, X. Zheng, S. Xiong, and A. Bao, 2023, Integration of UAV and GF-2 Optical Data for Estimating Aboveground Biomass in Spruce Plantations in Qinghai, China, *Sustainability (Switzerland)*, 15(12), doi: 10.3390/su15129700.
- [2] W. de O. S. Ana et al., 2022, Methodology for Defining Priority Locations for Carrying out a Forest Inventory in Points with Conflict between Urban Tree Planting and the Electricity Grid, *Energies (Basel)*, 15(3), doi: 10.3390/en15030684.
- [3] P. R. Rajbangshi and D. Nambiar, 2020, who will stand up for us? the social determinants of health of women tea plantation workers in India, *Int J Equity Health*, 19(1), doi: 10.1186/s12939-020-1147-3.
- [4] Y. Chen and S. Tian, 2020, Feature-level fusion between gaofen-5 and sentinel-1a data for tea plantation mapping, *Forests*, 11(12), pp. 1–21, doi: 10.3390/f11121357.
- [5] J. Kurihara, V. C. Koo, C. W. Guey, Y. P. Lee, and H. Abidin, 2022, Early Detection of Basal Stem Rot Disease in Oil Palm Tree Using Unmanned Aerial Vehicle-Based Hyperspectral Imaging, *Remote Sens (Basel)*, 14(3), doi: 10.3390/rs14030799.
- [6] A. R. Cameron, A. Meyer, C. Faverjon, and C. Mackenzie, 2020, Quantification of the sensitivity of early detection surveillance, *Transbound Emerg Dis*, 67(6), pp. 2532–2543, doi: 10.1111/tbed.13598.
- [7] A. Afrizal, O. Hospes, W. Berenschot, A. Dhialulhaq, R. Adriana, and E. Poetry, 2023, Unequal access to justice: an evaluation of RSPO's capacity to resolve palm oil conflicts in Indonesia, *Agric Human Values*, 40(1), pp. 291–304, doi: 10.1007/s10460-022-10360-z.
- [8] J. Wang et al., 2023, Recognition of the Interaction Mechanisms between Water and Land Resources Based on an Improved Distributed Hydrological Model, *Water (Switzerland)*, 15(10), doi: 10.3390/w15101946.
- [9] N. H. Buitendijk et al., 2022, More grazing, more damage? Assessed yield loss on agricultural grassland relates nonlinearly to goose grazing pressure, *Journal of Applied Ecology*, vol. 59, no. 12, pp. 2878–2889, doi: 10.1111/1365-2664.14306.
- [10] C. Luo, B. Zhang, J. Liu, X. Wang, F. Han, and J. Zhou, 2020, Effects of different ages of Robinia pseudoacacia plantations on soil physiochemical properties and microbial communities, *Sustainability (Switzerland)*, 12(21), pp. 1–18, doi: 10.3390/su12219161.
- [11] W. Xiang et al., 2022, Spatial Representativeness of Eddy Covariance Measurements in a Coniferous Plantation Mixed with Cropland in Southeastern China, *Remote Sens (Basel)*, 14(19), doi: 10.3390/rs14195022.
- [12] K. Moodley, P. V. Hernandez-Serrano, A. J. Zaveri, M. G. H. Schaper, M. Dumontier, and G. Van Dijck, 2020, The Case for a Linked Data Research Engine for Legal Scholars, *European Journal of Risk Regulation*, 11(1), pp. 70–93, doi: 10.1017/err.2019.51.
- [13] N. Berggren and C. Bjørnskov, 2023, Does legal freedom satisfy, *European Journal of Law and Economics*, 55(1), pp. 1–28, doi: 10.1007/s10657-022-09753-6.
- [14] D. Braun, 2023, I beg to differ: how disagreement is handled in the annotation of legal machine learning data sets, *Artif Intell Law (Dordr)*, doi: 10.1007/s10506-023-09369-4.
- [15] G. Bates, I. Cinar, and M. Nalepa, 2019, Award from the Political Organizations and Parties section of APSA. She has published her research in the, *Studies in Grammar, Logic and Rhetoric, and Decyzje. Association*, 18(1), p. 161, doi: 10.7910/DVN/1HCPSG.
- [16] I. Numata, A. J. Elmore, M. A. Cochrane, C. Wang, J. Zhao, and X. Zhang, 2022, Deforestation, plantation-related land cover dynamics and oil palm age-structure change during 1990-2020 in Riau Province, Indonesia, *Environmental Research Letters*, 17(9), doi: 10.1088/1748-9326/ac8a61.
- [17] P. Maring, “The strategy of shifting cultivators in West Kalimantan in adapting to the market economy: empirical evidence behind gaps in interdisciplinary communication.”
- [18] A. Fitch et al., 2022, The Coffee Compromise: Is Agricultural Expansion into Tree Plantations a Sustainable Option?,” *Sustainability (Switzerland)*, 14(5), doi: 10.3390/su14053019.
- [19] H. Lan, K. Stewart, Z. Sha, Y. Xie, and S. Chang, 2022, Data Gap Filling Using Cloud-Based Distributed Markov Chain Cellular Automata Framework for Land Use and Land Cover Change Analysis: Inner Mongolia as a Case Study, *Remote Sens (Basel)*, 14(3), doi: 10.3390/rs14030445.
- [20] D. Braun, 2023, I beg to differ: how disagreement is handled in the annotation of legal machine learning data sets, *Artif Intell Law (Dordr)*, doi: 10.1007/s10506-023-09369-4.
- [21] “fs20_A mixed methods approach for measuring topic sensitivity in conservation”.
- [22] N. Segura-Alabart, F. Serratos, S. Gómez, and A. Fernández, 2022, Nonunique UPGMA clusterings of microsatellite markers, *Briefings in Bioinformatics*, 23(5), doi: 10.1093/bib/bbac312.
- [23] Y. Lin and S. Chen, 2020, A Centroid Auto-Fused Hierarchical Fuzzy c-Means Clustering, [Online]. Available: <http://arxiv.org/abs/2004.12756>
- [24] J. Anibal et al., 2022, HAL-X: Scalable hierarchical clustering for rapid and tunable single-cell analysis, *PLoS Comput Biol*, 18(10), doi: 10.1371/journal.pcbi.1010349.
- [25] Z. Liu, Y. Li, L. Yao, X. Wang, and F. Nie, 2021, Agglomerative Neural Networks for Multi-View Clustering.
- [26] A. A. Shastri, K. Ahuja, M. B. Ratnaparkhe, and Y. Busnel, 2021, Probabilistically sampled and spectrally clustered plant species using phenotypic characteristics, *PeerJ*, 9, doi: 10.7717/peerj.11927.
- [27] J. Xu et al., 2021, Effects of randomized management on the forest distribution patterns of Larix Kaempferi plantation in Xiaolongshan, Gansu province, China, *Forests*, 12(8), doi: 10.3390/f12080981