

# Klasterisasi Peraturan Daerah di Seluruh Wilayah Jawa dengan Menggunakan Algoritma K-means

Nur Aini Rakhmawati<sup>\*[1]</sup>, Deanda Bevani Aletha<sup>[2]</sup>, Luh Putu Gayatri Widiastuti<sup>[3]</sup>, Alexander Sidum Laka Kaki<sup>[4]</sup>, Achmad Thoriq Aminullah<sup>[5]</sup>

Sistem Informasi, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia  
nur.aini@is.its.ac.id<sup>[1]</sup>, deandabvn@gmail.com<sup>[2]</sup>, gayatri.putri@gmail.com<sup>[3]</sup>, alexsidum@gmail.com<sup>[4]</sup>, dantethoriq80@gmail.com<sup>[5]</sup>

**Abstract**— The Local Indonesian government generates various legal products. This study aims to cluster the titles of regional regulations in Java. The data are collected from law repositories of all provincial governments in Java. The repository is called as JDIH. The analytical method used is clustering with the K-means algorithm. We divide 1849 regulations into ten clusters. 890 data are in the cluster related to motor vehicle tax, public health and safety.

**Keywords**— JDIH, Clustering, K-means, legal product

**Abstrak**— Pemerintah Indonesia selama ini telah berupaya untuk menunjukkan kinerja yang dibuktikan dengan disahkannya berbagai produk hukum. Penelitian ini bertujuan untuk melakukan klasterisasi nama/judul peraturan daerah di Pulau Jawa. Data yang digunakan berupa semua peraturan pemerintah daerah mengenai berbagai aspek kehidupan bermasyarakat dan bernegara, yang berasal dari seluruh situs JDIH milik setiap pemerintah daerah provinsi di Pulau Jawa. Metode analisa yang digunakan adalah klasterisasi dengan algoritma K-means. Dari 1.849 peraturan daerah, kami dapat mengklusterisasi peraturan-peraturan tersebut menjadi 10 cluster. Dari keseluruhan data, 890 data berada pada klaster terkait peraturan daerah berbicara mengenai pajak kendaraan bermotor, kesehatan, dan keamanan masyarakat.

**Kata Kunci**— JDIH, Klasterisasi, K-means, produk hukum

## I. PENDAHULUAN

Negara Indonesia adalah negara hukum, jelas dinyatakan dalam UUD 1945 pasal 1 ayat (3) [1]. Begitu banyak produk-produk hukum yang telah dibuat selama masa pemerintahan berjalan hingga sampai saat ini. Dalam menunjang negara hukum yang berdaulat, daerah diberikan hak otonomi daerah. Hak otonomi daerah ini memungkinkan pemerintah daerah memiliki wewenang untuk mengatur urusan pemerintahannya secara mandiri. Dengan hak tersebut, seluruh daerah di seluruh Indonesia bisa membuat dan merumuskan kebijakan daerahnya sendiri terkait peningkatan layanan masyarakat maupun pemberdayaan masyarakat lokal. Kebijakan itulah yang nantinya menginisiasi dan melahirkan pembentukan

peraturan daerah.

Menurut data yang dimuat di halaman awal situs jdih.go.id, ada sekitar 155.097 produk daerah yang terdaftar dan masuk JDIHN (Jaringan Dokumentasi dan Informasi Hukum Nasional). Produk daerah ini terbagi menjadi Peraturan Daerah Tingkat Provinsi/Kabupaten/Kota, Peraturan DPRD Tingkat Provinsi/Kabupaten, Peraturan Gubernur/Bupati/Walikota, dll. Jaringan Dokumentasi dan Informasi Hukum (JDIH) memberikan informasi berupa kumpulan peraturan dan kebijakan yang dibuat dan/atau diberlakukan di wilayah terkait. Informasi yang diperoleh dari JDIH suatu daerah mampu mengatur atau setidaknya menggambarkan tata cara hidup kelompok masyarakat di daerah tersebut. Penelitian yang berhasil kami temukan melakukan analisis terhadap bagaimana dampak implementasi teknologi informasi pada JDIH daerah Semarang dalam fungsi mengatur proses dan pengambilan keputusan terkait legislasi di daerah Semarang [2].

Dari sudut pandang teknologi, saat ini sudah banyak berkembang tools atau alat bantu untuk mendapatkan informasi hasil olahan dari data berukuran besar. Salah satunya adalah library sklearn atau scikit-learn, library ini menggunakan bahasa python yang bersifat open source, library ini cukup populer dan banyak digunakan sehingga perkembangan dan kompetensinya dapat dipertanggung jawabkan. Oleh karena itu, kami menilai bahwa library ini layak untuk digunakan sebagai metode klasterisasi pada penelitian ini.

Untuk melakukan klasterisasi atau pengelompokan, kami membutuhkan sebuah cara yang dapat secara otomatis mengolah data peraturan daerah di seluruh Jawa. Muncul sebuah penelitian yang berhasil menggunakan algoritma K-means untuk mengelompokkan daerah aliran sungai. Penelitian tersebut mengklasterisasi sungai berdasarkan tingkat kekritisannya untuk diprioritaskan dalam upaya rehabilitasi daerah aliran sungai [3].

Dengan mempertimbangkan hasil dua penelitian yang telah dijabarkan sebelumnya, penulis terdorong untuk melakukan analisis mengenai aspek yang diatur dalam produk hukum pemerintah daerah di seluruh Jawa. Penelitian ini

dilakukan untuk mengetahui objek/subjek yang paling banyak diikat peraturan pemerintah daerah, dengan mengkategorikan semua judul peraturan daerah berdasarkan angka kemunculan (modus). Penelitian ini bermanfaat untuk menganalisis peraturan terkait di setiap daerah sehingga bisa diketahui di setiap cluster terkait membahas seputar topik apa dan peraturan di daerah mana di Pulau Jawa yang membahas hal terkait tersebut.

## II. TINJAUAN PUSTAKA

### A. ParseHub

Menurut Anand V. Saukar (2018), ParseHub merupakan perangkat lunak untuk komputer, yang sederhana dan berfungsi untuk mengekstrak banyak data dari suatu halaman web [4]. Singkatnya, ParseHub merupakan suatu tools web scraping, Web Scraping adalah proses pengambilan sebuah dokumen semi - terstruktur dari internet, umumnya berupa halaman - halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain [5].

### B. Clustering

Analisis kluster adalah aktivitas mengidentifikasi pola (biasanya diwakili sebagai vektor pengukuran, atau titik dalam ruang multidimensi) dan mengelompokkannya ke dalam cluster berdasarkan kategori tertentu [6].

Analisis kluster membagi data menjadi beberapa kluster (kelompok) yang memiliki arti, berguna, atau keduanya. Untuk mendapatkan pengkategorian dari hasil pencarian yang baik, maka dapat diterapkan beberapa algoritma yang ada pada klusterisasi, salah satu diantaranya yaitu Suffix Tree Clustering (STC), K-Means, dan algoritma lainnya [7].

### C. K-Means

K-means adalah salah satu algoritma yang digunakan untuk klusterisasi, tujuan dari k-means sendiri adalah untuk membagi data masukan menjadi beberapa kelompok. Pada algoritma k-means ini, komputer mengelompokkan dengan sendiri data-data yang telah terinput tanpa perlu mengetahui terlebih dulu tujuan kelasnya, oleh karena itu k-means termasuk dalam unsupervised learning [8].

Masukkan yang diterima dalam algoritma ini adalah data atau objek dan k buah kelompok/cluster yang diinginkan, yang selanjutnya dengan algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Dan pada setiap kelompok tersebut terdapat titik pusat atau biasa disebut centroid yang berfungsi sebagai representasi kelompok tersebut.

### D. Scikit-Learn

Scikit-learn atau yang biasa dikenal dengan sklearn adalah sebuah perangkat lunak open source yang biasa digunakan untuk machine learning dalam bahasa pemrograman Python [9], library ini mendukung berbagai macam algoritma seperti klasifikasi, regresi, dan clustering, termasuk algoritma k-means, scikit-learn juga merupakan salah satu library yang

paling populer di github [10].

### E. JDIH

Berdasarkan Peraturan Presiden Nomor 33 Tahun 2012, JDIH merupakan media/sarana untuk memuat, mengakses, mengelola, dan pelayanan dokumen dan informasi hukum yang berintegritas, mudah diakses, cepat, dan tepat. Dokumen dan informasi yang dimuat di dalam JDIH berupa beberapa peraturan perundang-undangan atau produk hukum [11].

## III. METODOLOGI

Pengumpulan data dilakukan dengan menggunakan scraping tool 'ParseHub'. Data yang dikumpulkan adalah judul peraturan daerah dari situs JDIH milik pemerintah daerah provinsi di seluruh wilayah Jawa. Setelah itu, pemilihan data mana saja yang hendak diekstrak dilakukan. ParseHub sendiri memiliki fitur yang akan secara otomatis menebak data lain yang serupa dengan yang diminta, dari halaman lain pada website tersebut. Setelah selesai memilih semua data, proses scraping (get data) dilakukan. Apabila semua data sudah selesai dipanen, maka selanjutnya dataset web scraping akan disimpan dalam format CSV di alamat lokal komputer.

Setelah dataset berhasil disimpan, klusterisasi data set dilakukan dengan menggunakan k-means [7]. Klusterisasi memanfaatkan library yang tersedia untuk python yaitu sklearn, dan 2 tools yang ada didalam library tersebut, untuk mengekstrak data dalam bentuk text kami menggunakan tools TfidfVectorizer. Proses pembuatan model kluster dilakukan secara berulang hingga menemukan jumlah kluster yang optimal berdasarkan nilai error yang muncul pada setiap iterasi.

Berikut adalah langkah -langkah pengambilan dan pengolahan data dari JDIH di Pulau Jawa yang digambarkan seperti pada Gambar 1.

Dalam melakukan proses pengambilan data dimulai dengan teknik scraping pada website JDIH.go.id untuk mengambil dokumen peraturan daerah yang ada di Pulau Jawa menggunakan tools web scraping bernama Parsehub. Di dalam Parsehub ini dilakukan pengecekan terkait data yang telah terekstrak sehingga kita bisa mengetahui data mengenai dokumen apa saja yang memungkinkan terlewatkan. Setelah data terekstrak semua, digunakan algoritma K-means yang tersedia [12] untuk mengklusterisasi data untuk dikelompokkan menjadi beberapa cluster berdasarkan kesamaan kata yang terkandung di dalam peraturan daerah tersebut. Dari hasil cluster tersebut akan divisualisasikan ke dalam bentuk grafik batang untuk menarik kesimpulan dari hasil clusterisasi tersebut.



Gambar 1: Metodologi Penelitian

#### IV. ANALISA

Dari proses scraping dan didapatkan sebanyak 1.849 judul peraturan melalui JDIH untuk seluruh daerah di Pulau Jawa, dilakukan klasterisasi dengan metode K-means menggunakan program yang telah disusun dengan bantuan library sklearn [13], dan dari hasil program tersebut dipilih hasil dengan nilai error yang terkecil sebagai hasil akhirnya pada Tabel 1 yang menunjukkan hasil clustering dan Tabel 2 yang menunjukkan jumlah data setiap provinsi pada cluster.

TABLE I. HASIL CLUSTERING DENGAN ALGORITMA K-MEANS

Cluster	Tentang	Kata Teratas	Jumlah
Cluster 0	pendirian, perusahaan, dan modal	Perusahaan; Bentuk; Pendirian; Hukum; Modal; Pasar; Penyertaan; Kredit; Jaya; Badan	85

Cluster	Tentang	Kata Teratas	Jumlah
Cluster 1	pajak kendaraan bermotor, kesehatan, dan keamanan masyarakat	Pajak; Retribusi; Modal; Pemerintah; Kendaraan; Pelayanan; Penyertaan; Bermotor; Kesehatan; Umum	890
Cluster 2	pembentukan dan tata kerja organisasi	Pembentukan; Susunan; Organisasi; Tata; Kerjasama; Perangkat; Kerja; Badan; Kantor; Pertanian	100
Cluster 3	anggaran belanja dan perhitungan pertanggungjawaban	Anggaran; Belanja; Pendapatan; Perhitungan; Sisa; Pertanggungjawaban; Pelaksanaan; Anggaran; Pendapoatan; Tingkat	244
Cluster 4	usaha, perizinan, dan pemberdayaan	Usaha; Milik; Badan; Kecil; Perizinan; Retribusi; Perikanan; Mikro; Menengah; Pemberdayaan	64
Cluster 5	pengelolaan, keuangan, dan pemanfaatan	Pengelolaan; Air; Barang; Keuangan; Milik; Tanah; Pokok; Pengambilan; Permukaan; Pemanfaatan	121
Cluster 6	penyelenggaraan, pendidikan, dan perlindungan	Penyelenggaraan; Perhubungan; Pendidikan; Kearsipan; Kesehatan; Keolahragaan; Ketenagakerjaan; Perlindungan; Kependudukan; Administrasi	134
Cluster 7	retribusi, biaya, dan pemerintahan	Praja; Pasar; Tarip; Penjelasan; Bea; Penjualan; Retribusi; Biaya; Tambahan; Pemerintah	64
Cluster 8	pembangunan, jangka, dan menengah	Pembangunan; Rencana; Jangka; Menengah; Dasar; Pola; Panjang; Induk; Kepariwisataaan; Tbk	83

Cluster	Tentang	Kata Teratas	Jumlah
Cluster 9	lembaga dan pembentukan	Organisasi; Tata; Kerja; Sekretariat; Lembaga; Dewan; Perwakilan; Kerjanas; Pembentukan; Susunan	64

Clust er	Jumlah						Provinsi Mayorit as
	Jati m	Jaten g	DI Y	Jaba r	DK I	Bante n	
4	14	6	17	13	9	5	DIY
5	31	21	15	30	10	14	Jatim
6	20	28	24	36	4	22	Jabar
7	4	0	58	1	1	0	DIY
8	20	13	14	17	9	10	Jatim
9	20	7	16	14	3	4	Jatim
<b>Total</b>	502	186	566	276	182	137	1849

Pada Tabel 1 didapatkan hasil pengelompokan data berdasarkan kemiripan kata yang terkandung di dalam peraturan daerah yang tersebar di seluruh Pulau Jawa. Dari algoritma K-means dihasilkan 10 kata yang paling sering muncul dan memiliki kemiripan antara cluster lainnya. Dari 10 kata yang sering muncul tersebut kami melakukan penentuan pada setiap cluster tersebut apa yang menjadi pembahasan di dalam peraturan daerah yang ada.

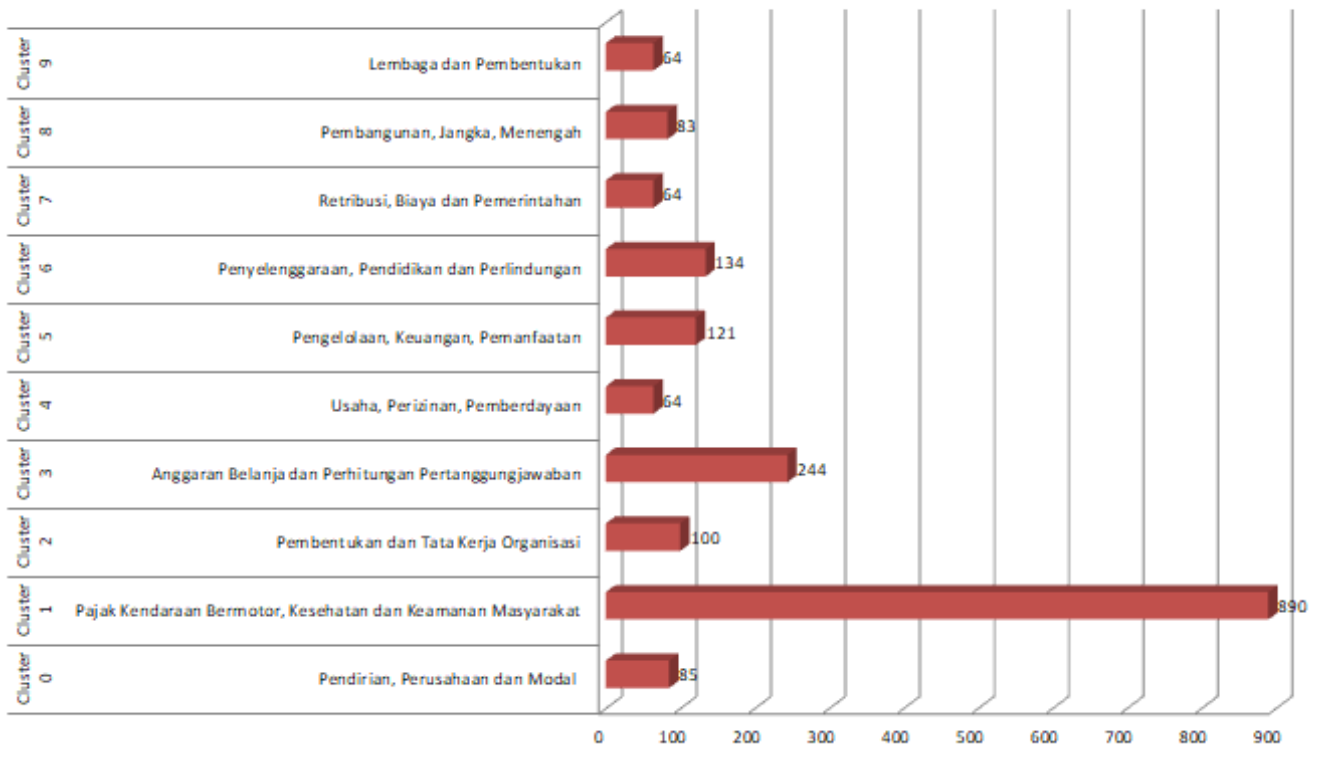
Pada Tabel 2 dengan menggunakan algoritma K-means , kami bisa menganalisis dari 1849 jumlah data hasil clustering di setiap Provinsi yang ada di Pulau Jawa. Dari mayoritas persebaran data tersebut, kami mengelompokan di setiap cluster provinsi yang paling banya jumlah data yang ada dan terdaftar di dalam website JDIH.go.id. Pada Tabel tersebut terlihat provinsi Jawa Timur mendominasi dalam menghasilkan peraturan.

TABLE II. JUMLAH DATA SETIAP PROVINSI PADA CLUSTER

Clust er	Jumlah						Provinsi Mayorit as
	Jati m	Jaten g	DI Y	Jaba r	DK I	Bante n	
0	29	9	15	16	11	5	Jatim
1	279	86	237	138	84	66	Jatim
2	13	4	65	6	6	6	DIY
3	72	12	105	5	45	5	DIY

Gambar 2 menampilkan kluster dan pembahasan tentang apa yang ada di dalam peraturan daerah di Pulau Jawa dengan menggunakan histogram. Pada grafik tersebut kluster paling banyak adalah kluster yang membahas pajak kendaraan, kesehatan dan keamanan.

### Clustering Peraturan Daerah



Gambar 2. Grafik Clustering Peraturan Daerah

## V. KESIMPULAN

Berdasarkan analisa yang dilakukan dapat disimpulkan bahwa kata yang dikandung di dalam produk peraturan daerah di Jawa beraneka ragam di sektor bidang yang berbeda pula di setiap peraturan. Perbedaan tersebut diklusterisasi berdasarkan persamaan kata yang dikandung. Dari 1.849 peraturan daerah, kami dapat mengklusterisasi peraturan-peraturan tersebut menjadi 10 cluster. Cluster 0 membahas mengenai pendirian, perusahaan dan modal dengan provinsi Cluster 1 membahas mengenai pajak kendaraan bermotor, kesehatan dan keamanan masyarakat. Cluster 2 membahas mengenai pembentukan dan tata kerja organisasi. Cluster 3 membahas mengenai anggaran belanja dan perhitungan pertanggungjawaban. Cluster 4 membahas mengenai usaha, perizinan dan pemberdayaan. Cluster 5 membahas mengenai pengelolaan, keuangan dan pemanfaatan. Cluster 6 membahas mengenai penyelenggaraan, pendidikan dan perlindungan. Cluster 7 membahas mengenai retribusi, biaya dan pemerintahan. Cluster 8 membahas mengenai pembangunan jangka menengah. Cluster 9 membahas mengenai lembaga dan pembentukan. Dari ke-10 cluster ini, dapat dilihat bahwa kata yang sering muncul adalah kata yang berhubungan dengan "Pembentukan" dan "Pendirian" yang menandakan bahwa peraturan daerah di Jawa banyak membahas mengenai regulasi dan hukum mengenai pembentukan dan pendirian terhadap sesuatu, seperti organisasi, badan, perusahaan, dan sebagainya. Dari keseluruhan data, cluster 1 memiliki jumlah data yang paling banyak yaitu sebesar 890 data yang menandakan kebanyakan pembahasan peraturan daerah berbicara mengenai "pajak kendaraan bermotor, kesehatan, dan keamanan masyarakat".

## REFERENCES

- [1] Undang-Undang Dasar Negara Republik Indonesia 1945 Pasal 1 Ayat (3).
- [2] Salami, M. and Kurniawan, A.T., 2016. Analisis Peran Jaringan Dokumentasi dan Informasi Hukum (JDIH) dalam Mendukung Proses Legislasi Daerah pada Bagian Hukum Sekretariat Daerah Kabupaten Semarang. *Jurnal Ilmu Perpustakaan*, 5(4), pp.91-100.
- [3] Turland, M. 2010. *php architect's Guide to Web Scraping with PHP*. Introduction-Web Scraping, str. 2.
- [4] Saukar, Anand V. (2018). An Overview On Web Scraping Techniques And Tools. *International Journal on Future Revolution in Computer Science & Communication Engineering* ISSN: 2454-4248 Volume: 4 Issue: 4.
- [5] Sudiro, Fauziah Putri, and S. T. P. Murtiningrum. Pengelompokan Daerah Irigasi di Kabupaten Sleman Berdasarkan Indeks Kinerja Daerah Irigasi Menurut Peraturan Menteri Pu Nomor 32/Prt/M/2007 Dengan Metode Fuzzy Clustering. Diss. [Yogyakarta]: Universitas Gadjah Mada, 2014.
- [6] McLachlan, G.J. and Basford, K.E., 1988. ; (Vol. 38). New York: M. Dekker.
- [7] Langgeni, D.P., Baizal, Z.A. and AW, Y.F., 2015, July. Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection. In *Seminar Nasional Informatika (SEMNASIF)* (Vol. 1, No. 4).
- [8] X. Wu and V. Kumar, eds., *The Top Ten Algorithms in Data Mining*. Chapman and Hall, 2009.
- [9] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [10] "The State of the Octoverse: machine learning". The GitHub Blog. GitHub. 2019-01-24. Retrieved 2019-10-17.
- [11] Pemerintah Indonesia. 2012. Peraturan Presiden Republik Indonesia Nomor 33 Tahun 2012 tentang Jaringan Dokumentasi dan Informasi Hukum Nasional. Lembaran Negara Republik Indonesia Tahun 2012 Nomor 82. Jakarta : Sekretariat Negara.
- [12] Nuraini Rakhmawati, Achmad Thoriq, Luh Putu Gayatri Wirastuti, Deanda Bevani Aletha, & Alexander Sidum Laka Kaki. (2020, March 1). Data Peraturan Pulau Jawa. <https://doi.org/10.5281/zenodo.3700461>
- [13] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *Journal of machine learning research* 12.Oct (2011): 2825-2830.