



Algoritma *K-Means* Untuk Mengklasifikasi Tingkat Layanan Informasi Hukum di Provinsi Jawa Barat

Beny Ruhiman^{1,2}, Ade Ramdan^{1,3} dan Christina Juliane^{1,4}

¹STMIK LIKMI Bandung, Magister Sistem Informasi

²Provinsi Jawa Barat, Biro Hukum dan HAM, email: beny.ruhiman@jabarprov.go.id

³Badan Riset dan Inovasi Nasional (BRIN), email: ader004@brin.go.id

⁴STMIK LIKMI Bandung, email: christina.juliane@likmi.ac.id

[1] Abstrak

Jaringan Dokumentasi dan Informasi Hukum (JDIH) adalah salah satu instansi pemerintah dalam bidang hukum, yang diatur melalui Peraturan Presiden Nomor 33 Tahun 2012. JDIH terdiri dari pusat dan daerah, pada tingkat provinsi Biro Hukum dan HAM adalah pusat JDIH di wilayahnya. JDIH di tingkat provinsi mempunyai tugas dan fungsi untuk melakukan pembinaan, monitoring dan evaluasi anggotanya. Untuk mengukur tingkat kepuasan masyarakat anggota JDIH di Jawa Barat, dilakukan *survey* dengan menggunakan metode *review 360* melalui website <https://jdih.jabarprov.go.id/review.php/>. Hasil *survey* terdapat data mentah sebanyak 18.045. Setelah dilakukan pra-pemrosesan dihasilkan 46 dataset dan 11 atribut. Pemodelan cluster menggunakan algoritma *K-Means*, hasil dari *cluster* data dievaluasi dengan metode *Davies Boulding Index (DBI)*. Hasil evaluasi menunjukkan tingkat kemiripan yang rendah sehingga jarak antar *cluster* semakin tinggi. Analisis hasil dari penelitian ini terklasifikasi 4 kluster, indikator kepuasan terendah diketahui pada *cluster* ke-3 yang terdiri dari 10 daerah. Hasil penelitian ini dapat menjadi bahan masukan bagi Pemerintah Provinsi Jawa Barat untuk menentukan kebijakan.

Kata kunci: *kepuasan layanan, review 360, algoritma k-means, Davies Boulding Index*

[2] Abstract

Legal Documentation and Information Network (JDIH) is one of the government agencies in the field of law, which is regulated by Presidential Regulation Number 33 of 2012. JDIH consists of central and regional levels, at the provincial level the Bureau of Law and Human Rights is the center of JDIH in its territory. JDIH at the provincial level has the duties and functions to provide guidance and evaluation at the members. To measure the level of community satisfaction with JDIH members in West Java, a survey was conducted using the 360 review method on the website <https://jdih.jabarprov.go.id/review.php/>. The results contained 18.045 raw data. After preprocessing, 46 datasets and 11 attributes were generated. Cluster modeling uses the K-Means algorithm, the results are evaluated by the Davies Boulding Index (DBI) method. Evaluation results show a low level of similarity so that the distance between clusters is getting higher. On this study is classified into 4 clusters, the lowest satisfaction indicator is known to be in cluster 3 which consists of 10 regions. In this research can be used to determining policies for the government of West Java Province.

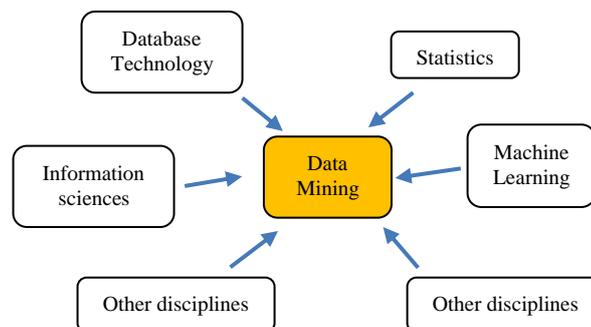
Keywords: *service satisfaction, review 360, algoritma k-means, Davies Boulding Index*

1. Pendahuluan

Dalam membangun budaya hukum diperlukan instrumen pendukung seperti perpustakaan dan dokumentasi hukum. Seminar ke III Hukum Nasional pada tahun 1974 dikemukakan perlu adanya kebijakan nasional dalam pembentukan Sistem Jaringan Dokumentasi dan Informasi Hukum (SJDIH). Namun secara formal, JDIH baru disahkan melalui Keputusan Presiden Nomor 91 Tahun 1999 (Lembaran Negara RI Tahun 1999 No.135). Seiring dengan perkembangan dinamika hukum dan juga teknologi informasi, peran JDIH diatur kembali melalui Peraturan Presiden Nomor 33 Tahun 2012 tentang Jaringan Dokumentasi dan Informasi Hukum Nasional. JDIH terdiri dari tingkat pusat dan daerah. Pada tingkat provinsi Biro Hukum dan HAM adalah pusat JDIH di wilayahnya. Tugas dan fungsinya yaitu melakukan pembinaan, monitoring dan evaluasi anggota.

Biro Hukum dan HAM Provinsi Jawa Barat memiliki anggota sebanyak 54 (lima puluh empat) terdiri dari (Sekretariat Daerah dan Sekretariat DPRD Kabupaten/Kota). Untuk mengukur kualitas layanan dilakukan *survey* dengan menggunakan metode 360 melalui *website* <https://jdih.jabarprov.go.id/review.php/>. *Survey* terdiri dari indikator-indikator yang ditetapkan melalui Keputusan Gubernur Jawa Barat 180/Kep.267-Hukham/2021 [1]. Indikator tersebut terdiri dari Organisasi, Sumberdaya Manusia, Teknis Pengelolaan, Koleksi Dokumen, Sarana dan Prasarana, Kepuasan Masyarakat, Laporan, Pemanfaatan TIK, Penguatan Organisasi dan Kerjasama serta Inovasi.

Data dari hasil *survey* tersebut perlu dilakukan penelitian, untuk mengetahui daerah yang kepuasan layanannya masih rendah agar dapat dilakukan pembinaan. Penelitian ini menggunakan teknik data *mining* yang merupakan bidang interdisipliner, pertemuan dari satu set disiplin ilmu, termasuk sistem basis data, statistik, pembelajaran mesin, visualisasi, dan ilmu informasi.



Gambar 1. Data Mining [1]

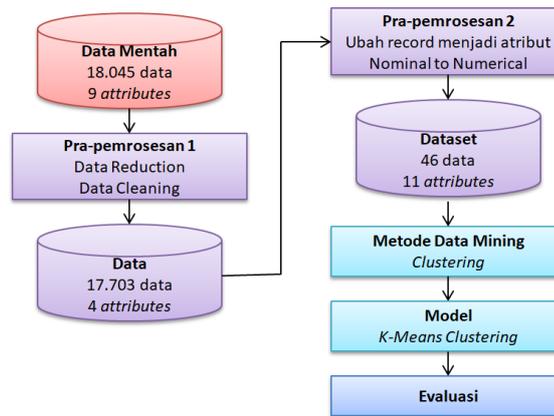
Dalam penggunaan metodologi pada data *mining* terdapat *performance issue* yang meliputi :

- *Efficiency and scalability of data mining algorithms* yaitu mengekstrak informasi secara efektif dari sejumlah besar data dalam basis data, algoritma data *mining* harus efisien dan terukur. Dengan kata lain, waktu berjalan dari algoritma data *mining* harus dapat diprediksi dan dapat diterima dalam basis data yang besar.
- *Parallel, distributed, and incremental mining algorithms*: besarnya ukuran basis data, distribusi data yang luas, dan kompleksitas komputasi dari beberapa metode data *mining* merupakan faktor yang memotivasi pengembangan algoritma data *mining* paralel dan terdistribusi. Algoritma tersebut melakukan modifikasi pengetahuan secara bertahap untuk mengubah dan memperkuat apa yang ditemukan sebelumnya [2].

Kontribusi dari makalah ini adalah sebagai berikut: pertama, kami melakukan ekstraksi data untuk mendapatkan suatu pola penentuan daerah-daerah yang kurang bagus dengan *K-Means Clustering*. Kedua, data yang kami gunakan adalah hasil pengumpulan data sendiri.

2. Metode Penelitian

Metode penelitian yang diusulkan adalah menggunakan metode data *mining clustering* untuk mengelompokkan daerah-daerah yang dihasilkan dari skor penilaian. Indikator yang digunakan adalah organisasi, sumberdaya manusia, teknis pengelolaan, koleksi dokumen, sarana dan prasarana, kepuasan masyarakat, laporan, pemanfaatan teknologi komunikasi dan informatika, penguatan organisasi, kerjasama dan inovasi. *Clustering* yang dilakukan dalam penelitian ini ditunjukkan pada gambar 2.



Gambar 2. Alur Penelitian

Data mentah (*raw data*) merupakan data yang didapat dari suatu sumber data dan informasi yang belum diolah maupun di analisa. Data mentah dalam penelitian ini diambil dari hasil *survey* melalui *website* yang dibuat oleh Biro Hukum dan HAM Sekretariat Daerah Provinsi Jawa Barat melalui <https://jdih.jabarprov.go.id/review.php>, dihasilkan *raw data* sebanyak 18.045 seperti yang ditunjukkan pada gambar 3. Data mentah kemudian dilanjutkan dengan pra-pemrosesan data untuk mengubah data mentah menjadi data yang dapat diproses oleh data *mining*.

id_nilai	kategori	nama_kab	nama_unit	indikator	nm_pertany...	nilai	tahun	tgl_update
37958	dprd	Kabupaten Ci...	jdihn	Organisasi	Apakah orga...	73	2021	Sep 22, 2021...
37959	dprd	Kabupaten Ci...	jdihn	Sumberdaya Manusia	Apakah Sum...	70	2021	Sep 22, 2021...
37960	dprd	Kabupaten Ci...	jdihn	Koleksi Dokumen	Apakah kolek...	72	2021	Sep 22, 2021...
37961	dprd	Kabupaten Ci...	jdihn	Pemanfaatan TIK	Berikan penil...	70	2021	Sep 22, 2021...
37962	dprd	Kabupaten Ci...	jdihn	Penguatan Organisasi ...	Berikan penil...	73	2021	Sep 22, 2021...
37963	dprd	Kabupaten Ci...	jdihn	Sarana dan Prasarana	Apakah sara...	74	2021	Sep 22, 2021...
37964	dprd	Kabupaten Ci...	jdihn	Sosialisasi/Promosi	Berikan penil...	74	2021	Sep 22, 2021...
37965	dprd	Kabupaten Ci...	jdihn	Teknis Pengelolaan	Apakah tekni...	70	2021	Sep 22, 2021...
37966	dprd	Kabupaten Ci...	jdihn	Inovasi	Berikan nilai ...	73	2021	Sep 22, 2021...
37967	kabkota	Kabupaten Ci...	jdihn	Organisasi	Apakah orga...	74	2021	Sep 22, 2021...

Gambar 3. Data Mentah

Tahapan pra-pemrosesan sangat penting pada data *mining* karena kualitas dari input data sangat mempengaruhi kualitas *output* analisis yang dihasilkan [3]. Pra-pemrosesan yang dilakukan pada penelitian ini terdiri dari pra-pemrosesan ke-1 dan pra-pemrosesan ke-2.

Pra-pemrosesan ke-1 meliputi tahapan berikut ini:

1. Reduksi data yaitu untuk mengurangi atribut dengan membuang atribut-atribut yang tidak mengandung informasi yang berguna. Data hasil reduksi hanya menyertakan atribut yang diperlukan dalam proses data *mining* maka ukuran dari *dataset* menjadi kecil sehingga lebih efisien.
2. Pembersihan data yaitu membersihkan data yang tidak memiliki nilai (*missing value*) [4] dengan cara dihapus. Proses reduksi dan pembersihan menyebabkan jumlah data menjadi 17.703 data sampel yang terdiri dari 4 fitur.

Hasil dari pra-pemrosesan ke-1 ditunjukkan pada gambar 4.

kategori	nama_kab	indikator	nilai
kabkota	Kabupaten Purwakarta	Organisasi	93
kabkota	Kabupaten Purwakarta	Sumberdaya Manusia	90
kabkota	Kabupaten Purwakarta	Koleksi Dokumen	97
kabkota	Kabupaten Purwakarta	Pemanfaatan TIK	88
kabkota	Kabupaten Purwakarta	Penguatan Organisasi da...	86
kabkota	Kabupaten Purwakarta	Sarana dan Prasarana	86
kabkota	Kabupaten Purwakarta	Sosialisasi/Promosi	84
kabkota	Kabupaten Purwakarta	Teknis Pengelolaan	85
kabkota	Kabupaten Purwakarta	Inovasi	85
kabkota	Kota Depok	Organisasi	98

Gambar 4. Hasil pra-pemrosesan ke-1

Sedangkan pra-pemrosesan ke-2 terdiri dari tahapan berikut ini:

1. Merubah data atau *record* di atribut indikator diubah menjadi atribut. Proses ini dilakukan untuk mendapatkan fitur yang lebih banyak sehingga mempermudah dalam mengidentifikasi daerah-daerah berdasarkan nilai indikatornya. Akibat pengubahan data atau *record* indikator menjadi fitur tersebut menyebabkan data berkurang menjadi 46 *dataset* tetapi fitur bertambah menjadi 11 atribut.
2. Konversi data nominal menjadi numerik. Konversi tersebut diperlukan karena algoritma *K-Means* hanya bisa digunakan untuk *dataset* yang atributnya bernilai numerik.

Hasil dari pra-pemrosesan ke-2 ditunjukkan pada gambar 5.

nama_kab	Kepuasan M...	Organisasi	Sumberday...	Koleksi Dok...	Pemanfaata...	Penguatan ...	Sarana dan ...	Sosialisasi...	Teknis Peng...	Inovasi
0	94.200	76.580	74.840	77.370	76.680	77.260	76.420	76.950	75.050	70.210
1	92.810	83.150	81.400	81.500	78.400	81.350	82.150	81.100	81.750	77
2	86.280	83.220	81	82.110	82.110	79.720	79.110	80.500	80.500	77.170
3	99.520	83.590	80.650	77.820	81.940	80.530	81.240	82.410	81.180	80
4	95.210	84.060	83.060	83	84.180	83	82.650	79.120	83.530	71.650
5	89.300	86.060	83.530	86.180	85.760	79.120	79.820	77.760	83.760	74.350
6	98.590	85.500	82.610	84.220	84.390	83.220	84.670	78.670	83.780	79.610
7	94.630	83.110	80.370	80.630	78.370	80.210	78.320	77.680	80.160	74.050
8	93.200	85.410	82.880	84.820	87.240	84.590	86.940	84.350	85.060	74.350
9	93.750	83.290	80.590	77.760	78.820	79.180	78.180	77.410	80.350	69.120
10	93.270	76.100	74.500	75.300	76.450	71.450	75.500	73.450	75.800	70.200
11	92.610	77.700	72.800	74.550	73.550	73.300	73.850	71.900	73.100	69.400

Gambar 5. Hasil pra-pemrosesan ke-2

Setelah didapat *dataset* selanjutnya adalah menentukan metode data *mining* yang akan digunakan. Pemilihan metode data *mining* dilakukan untuk menghasilkan pola penting sebagai pengetahuan yang didapat dalam proses data *mining*. Metode data *mining* yang digunakan untuk mengelompokkan data indikator adalah Metode *Clustering*. *Clustering* adalah salah satu metode dalam data *mining* yang berfungsi untuk mengelompokkan data yang memiliki kemiripan karakteristik [5].

Tujuan dari *clustering* adalah mengelompokkan data-data yang hampir sama dalam satu area tertentu [6]. *Clustering* digunakan ketika kita tidak mengetahui bagaimana data harus dikelompokkan. Metode *clustering* yang salah satunya populer dan paling umum digunakan adalah algoritma *K-Means* [7], [8]. Algoritma *K-Means* tersebut yang akan digunakan dalam penelitian ini karena algoritma *K-Means* mudah dan sederhana saat diimplementasikan.

K-Means adalah salah satu algoritma *clustering* yang menggunakan metode *partitional clustering* [9]. Data *K-Means* dibagi ke dalam *cluster* yang terdiri dari data yang mirip dan berbeda karakteristiknya [9]. Sebuah *cluster* merupakan kumpulan data-data yang sama di antara mereka dan berbeda dengan data dari *cluster* lainnya. Jadi tujuan dari metode *K-Means Clustering* adalah mengelompokkan data berdasarkan kedekatan satu sama lainnya [10].

Tahapan terakhir adalah evaluasi terhadap kualitas *clustering* yaitu dengan menggunakan metode *Davies Boulding Index (DBI)*. *Davies Boulding Index* merupakan salah satu metode untuk mengevaluasi hasil algoritma *clustering*. *Davies Boulding Index* mengukur jarak antar *cluster*. Jika setiap data dalam *cluster* memiliki tingkat kemiripan yang rendah berarti jarak antar *cluster* tinggi. Sebaliknya, jika setiap data yang memiliki kemiripan karakteristik tinggi berarti jarak antar *cluster* akan rendah [11]. Jadi jumlah *cluster* terbaik ditunjukkan dengan nilai *Davies Boulding Index* yang semakin kecil [12].

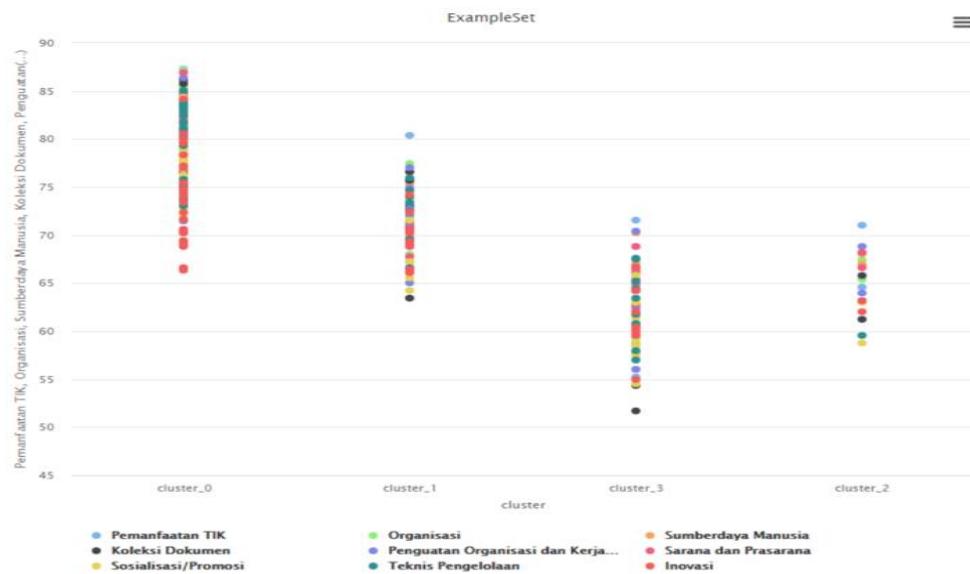
3. Hasil dan Pembahasan

Penelitian yang dilakukan membandingkan 8 (delapan) bentuk model *K-Means* dengan pemberian beberapa nilai parameter *input k* yang berbeda untuk menentukan *k optimum cluster*. *k* merupakan parameter input dari *K-Means*. Kemudian dilihat *k optimum* menggunakan nilai *Davies Boulding Index (DBI)* untuk *model K-Means* yang hasilnya ditunjukkan pada tabel 1.

Tabel 1. Pengukuran *k* dan nilai *Davies Boulding Index (DBI)*

K Klaster	Nilai <i>Davies Boulding Index</i>
3	0,100
4	0,079
5	0,095
6	0,085
7	0,088
8	0,080
9	0,083
10	0,092

Hasil penentuan *k optimum* untuk *K-Means* seperti yang ditunjukkan pada tabel 1, diperoleh *k optimum* berada pada saat *k = 4*. Parameter input *k = 4* merupakan *k optimum* karena memiliki nilai *Davies Boulding Index (DBI)* paling kecil yaitu 0,079 dibandingkan dengan nilai input *k* lainnya. *Cluster* yang dihasilkan dari algoritma *K-Means* dengan parameter *input k = 4* ditunjukkan pada gambar 6.

Gambar 6. Visualisasi hasil *clustering*

Hasil pemisahaan *dataset* menjadi 4 (empat) *cluster* tersebut merupakan pola pengelompokkan daerah-daerah berdasarkan skor nilai indikator. *Cluster* yang memiliki skor nilai terendah teridentifikasi dengan mudah dari 4 (empat) *cluster* tersebut. *Cluster* yang memiliki skor nilai yang terendah adalah *cluster* ke-3.

Tabel 2. Hasil *Clustering* pada *Cluster* ke 3

Nama Daerah	Indikator									
	KM	OG	SDM	KD	PT	POK	SaPras	SosPro	TP	IV
28	96.80	63.25	64.25	51.75	55.25	56.00	58.50	54.50	67.50	66.75
30	87.43	70.40	67.00	62.60	71.60	65.00	65.00	63.40	64.40	60.40
31	83.00	65.33	57.67	59.33	66.67	59.33	57.50	59.17	60.83	55.00
32	95.25	67.40	65.40	60.80	66.20	64.20	64.20	61.40	61.8	59.60
34	97.18	62.17	62.50	60.67	64.33	61.50	61.83	58.50	60.17	60.00
36	84.41	61.40	64.40	66.60	68.80	65.00	68.80	65.80	67.60	60.20
38	95.14	64.40	70.20	61.80	66.20	70.40	65.40	63.00	65.20	64.20
39	100.00	58.80	59.60	59.60	60.60	62.60	62.80	58.00	57.00	62.00
40	94.93	62.00	61.80	60.40	64.80	62.60	66.20	61.60	63.40	64.20
43	81.00	62.00	56.00	54.33	61.50	57.83	59.67	57.50	58.00	60.00

Keterangan Nama Daerah dengan nama inisial :
28 : DPRD Kabupaten Cr

Keterangan Indikator :
KM : Kepuasan Masyarakat
OG : Organisasi

30 : DPRD Kota Cn	SDM : Sumberdaya Manusia
31 : DPRD Kabupaten Bi	KD : Koleksi Dokumen
32 : DPRD Kota Dk	PT : Pemanfaatan TIK
34 : DPRD Kota Bi	POK : Penguatan Organisasi & Kerjasama
36 : DPRD Kota Br	SarPras : Sarana dan Prasarana
38 : DPRD Kabupaten Tsk	SosPro : Sasial dan Promosi
39 : DPRD Kota Ci	TP : Teknis Pengelolaan
40 : DPRD Kabupaten Kn	IV : Inovasi
43 : DPRD Kabupaten Ski	

Hasil *cluster* di tabel 2 menunjukkan terdapat satu kolom yang hasilnya berbeda yaitu tingginya kepuasan masyarakat. Hal tersebut disebabkan oleh jumlah dan hasil rata-rata skor penilaian oleh masyarakat umum. Namun karena *survey* menggunakan metode *review* 360, yang penilainya terdiri dari atasan, atasan langsung, sederajat, dan penilaian diri sendiri, maka dapat diketahui indikator lainnya masih rendah.

Hasil analisis pada *cluster* ke-3 merupakan *cluster* pada daerah yang memiliki tingkat kepuasan terendah. Tingkat kepuasan terendah diukur berdasarkan indikator organisasi, sumber daya manusia, koleksi dokumen, pemanfaatan TIK, penguatan organisasi, kerjasama, sarana dan prasarana, sosialisasi dan promosi, teknik pengelolaan serta inovasi. Hasil penelitian ini, dapat menjadi bahan masukan bagi Biro Hukum dan HAM Sekretariat Daerah Provinsi Jawa Barat untuk membuat kebijakan perbaikan kualitas layanan kepada daerah tersebut.

4. Kesimpulan

Hasil dari penelitian yang dilakukan dapat disimpulkan bahwa Algoritma *K-Means Clustering* dengan *Davies Boulding Index* dapat mengklasifikasi daerah berdasarkan skor nilai indikator dengan baik. Hasil *clustering* tersebut dapat diketahui daerah dengan skor terendah yaitu pada *cluster* ke-3. Terdapat *gap* perbedaan skor nilai indikator kepuasan masyarakat dibandingkan dengan indikator lainnya. Disarankan perlu perbaikan pada metode *survey* agar tidak terdapat *gap* yang terlalu jauh. Selain itu, perlu dilakukan pengujian menggunakan algoritma *clustering* lainnya.

[3] Daftar Pustaka

- [1] JDIH Provinsi Jawa Barat, Keputusan Gubernur Jawa Barat Nomor 180/Kep.267-Hukham/2021 tentang Indikator Penilaian Penghargaan Jaringan Dokumentasi dan Informasi Hukum Daerah Provinsi Jawa Barat. Indonesia, [Online]. Available: <http://jdih.jabarprov.go.id/assets/uploads/files/produk/2021kg00320267.pdf>, 2021. (peraturan)
- [2] Jiawei Han, "Data *mining*: Concepts and Techniques Second Edition", Elsevier, AID International, Sabre Foundation, 2006. (buku)
- [3] H. Juanedi, H. Budianto, I. Maryati, Y. Melani, "Data Transformation Pada Data *Mining*", 2011. (Prosiding Konferensi Nasional "Inovasi dalam Desain dan Teknologi" (IdeaTech))
- [4] A. Yang, W. Zhang, J. Wang, K. Yang, Y. Han, L. Zhang, "Review on the Application of Machine Learning Algorithms in the Sequence Data *Mining* of DNA", *Frontiers in Bioengineering and Biotechnology*, 2020. (jurnal)

- [5] G. Biswas; J.B. Weinberg; D.H. Fisher, "ITERATE: a conceptual clustering algorithm for data *mining*" IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Volume: 28, Issue: 2, May 1998. (jurnal)A. Bastian., "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)", Jurnal Sistem Informasi, 2018. (jurnal)
- [6] Raykov, Y. P., Boukouvalas, A., Baig, F., and Little, M. A., "What to do when K-means clustering fails: a simple yet principled alternative algorithm", PloS one, vol. 11, no. 9, 2016. (jurnal)
- [7] I. Budiman, T. Prahasto, and Y. Christyono, "Data Clustering Menggunakan Metodologi Crisp-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma", presented in 2012 Seminar Nasional Aplikasi Teknologi Informasi (SNATI 2012), Yogyakarta. (seminar)
- [8] Sitompul, B.J.D., Sitompul, O.S., Sihombing, P., "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algoritma", The 3rd International Conference on Computing and Applied Informatics 2018, (konferensi).
- [9] Nawrin, S., Rahman, M.R. & Akhter, S., "Exploreing K-Means with internal validity indexes for data clustering in traffic management system", International Journal of Advanced Computer Science and Applications, 2017.(jurnal)
- [10] Arora Pr., Deepali Dr., Varshney Sh., "Analysis of K-Means and K-Medoids Algorithm For Big Data", International Conference on Information Security & Privacy (ICISP 2015). (konferensi)
- [11] Thakare, Y.S. & Bagal, S.B., "Performance evaluation of k-means clustering algorithm with various distance metrics", International Journal of Computer Applications, 2015. (jurnal)
- [12] Davies, D. L., Bouldin, D. W. "A Cluster Separation Measure", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1979.(jurnal)