

Perbandingan Kinerja Algoritma K-Means dan K-Medoids Dalam Klasterisasi Jumlah Tindak Pidana Kejahatan Berbasis Wilayah Kepolisian Daerah

Gelar Nurcahya¹, Arief Wibowo², Dwi Kristanto³

¹Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur
Jl. Ciledug Raya, DKI Jakarta, Indonesia

e-mail: gelarnurcahya@gmail.com¹, arief.wibowo@budiluhur.ac.id², dwi.kristanto@budiluhur.ac.id³

Received : November, 2023

Accepted : November, 2023

Published : Desember, 2023

Abstract

Criminal acts are often a problem that occurs in Indonesia. Where currently the number of reports handled by the police regarding criminal acts is always there every day. Indonesia's population is increasing and the background of perpetrators who are unemployed is often one of the reasons why the police find it difficult to resolve criminal acts that occur due to limited human resources. To overcome this problem, information is needed that provides areas in Indonesia where criminal acts frequently occur so that the police can make decisions to allocate human resources to protect those jurisdictions from criminal acts that occur. Using data on criminal offenses and the employment of criminal offenders, namely not working from 2021, data was taken from the National Police Criminal Investigation Unit's Pusiknas Annual Journal. The data will be clustered using data mining techniques using the K-Means and K-Medoids algorithms. These 2 algorithms produced 2 clusters with the smallest Davies Bouldin index value found in the K-Means algorithm with a value of 0.272. With the research results which produced 2 clusters, it can be concluded that there are categories of high crime and low crime.

Keywords: Data Mining, K-Means, K-Medoids, klustering, Jurnal Pusiknas

Abstrak

Tindak pidana seringkali menjadi permasalahan yang terjadi di Indonesia. Dimana saat ini jumlah laporan yang ditangani pihak kepolisian terkait tindak pidana selalu ada setiap harinya. Jumlah penduduk Indonesia yang semakin meningkat dan latar belakang pelaku yang merupakan pengangguran seringkali menjadi salah satu penyebab kepolisian sulit menuntaskan tindak pidana yang terjadi karena keterbatasan sumber daya manusia. Untuk mengatasi permasalahan tersebut diperlukan informasi yang menyediakan wilayah-wilayah di Indonesia yang sering terjadi tindak pidana sehingga pihak kepolisian dapat mengambil keputusan untuk mengalokasikan sumber daya manusia untuk melindungi wilayah hukum tersebut dari tindak pidana yang terjadi. Menggunakan data tindak pidana dan pekerjaan pelaku pidana yaitu tidak bekerja mulai tahun 2021, data diambil dari Jurnal Tahunan Pusiknas Bareskrim Polri. Data tersebut akan dikelompokkan menggunakan teknik data mining dengan menggunakan algoritma K-Means dan K-Medoids. 2 algoritma ini menghasilkan 2 klaster dengan nilai indeks Davies Bouldin terkecil terdapat pada algoritma K-Means dengan nilai 0,272. Dengan hasil penelitian yang menghasilkan 2 klaster maka dapat disimpulkan terdapat kategori kejahatan tinggi dan kejahatan rendah.

Kata Kunci: Data Mining, K-Means, K-Medoids, klastering, Jurnal Pusiknas

1. PENDAHULUAN

Masyarakat dan pihak kepolisian sangat membutuhkan informasi mengenai intensitas angka kriminalitas yang terjadi di daerah tersebut [1]. Menurut data dari jurnal tahunan pusiknas bareskrim polri total tindak pidana yang terjadi di Indonesia sebanyak 274.988 pada tahun 2021 [2]. Dari data tindak pidana tersebut dapat dilakukan analisis dan pengelompokan data, pola yang di hasilkan dapat memberikan wawasan dan pengetahuan yang berharga bagi pihak kepolisian dalam merencanakan strategi pencegahan dan penanggulangan tindak pidana kejahatan. Menurut dataindonesia.id jumlah personel kepolisian di Indonesia berdasarkan laporan kepolisian RI jumlah personelnnya tercatat sebanyak 434.135 orang[3]. Maka dari itu akan digunakan Teknik *Data Mining* untuk dilakukan klastering, dengan mencari nilai DBI terbaik dengan menggunakan algoritma *K-Medoids* dan *K-Means*. *Data Mining* adalah suatu proses pencarian data yang tersembunyi dari database [4], [5]. Dimana sebelumnya terdapat penelitian yang meneliti mengenai tindak kejahatan menggunakan algoritma K-Means [1], [6]–[9] Algoritma yang dikenal sebagai klasifikasi K-Means mengelompokkan data menurut data yang paling dekat dengan titik pusat klaster [9]. K-Medoids adalah sebuah teknik klastering partisional yang memiliki tujuan untuk mengidentifikasi satu set *k-cluster* yang paling mewakili objek-objek dalam kumpulan data [10].

Alasan dilakukan penelitian ini adalah untuk mengidentifikasi pola dan kelompok-kelompok tindak pidana yang muncul dalam data kepolisian daerah, dengan harapan dapat memberikan wawasan baru dalam upaya pencegahan dan penanggulangan kejahatan serta mengefektifkan jumlah personel polri yang ada di Indonesia agar suatu wilayah tidak kekurangan anggota polisi untuk menindak tindak pidana yang terjadi. Dalam penelitian sebelumnya menggunakan data tindak pidana, algoritma yang digunakan adalah klasterisasi dengan K-Means [11]. dan ada pula penelitian yang mengelompokkan kasus kejahatan menggunakan K-Medoids [12]. Dengan penelitian yang dilakukan 2 jurnal tersebut maka pengujian model algoritma yang digunakan adalah K-Means dan K-Medoids untuk mendapatkan hasil terbaik dengan model algoritma tersebut, pengujian yang digunakan adalah Davies Bouldin Indeks (DBI) . Data yang di gunakan adalah data jumlah tindak pidana berdasarkan atribut-atribut tertentu jumlah tindak pidana berdasarkan tahun, tingkat pengangguran. Melalui penyelesaian prosedur *data mining* ini, wajar jika kita dapat mengenali contoh dan pola yang terkait dengan tindak pidana yang terjadi di wilayah Indonesia. Informasi yang digunakan dalam proses *data mining* tindak pidana diambil dari informasi jurnal pusiknas bareskrim polri seperti yang ditampilkan pada Gambar 1, Gambar 2.

NO	POLDA	JAN	FEB	MAR	APR	MEI	JUN	JUL	AGT	SPT	OKT	NOV	DES
1	Bareskrim Polri	54	85	76	71	51	51	50	70	71	55	45	60
2	Aceh	633	682	783	783	643	606	538	574	672	737	730	624
3	Sumut	3.157	3.231	3.330	3.096	2.996	3.378	3.170	3.083	3.380	3.473	3.802	3.720
4	Sumbang	427	548	560	773	509	583	529	510	525	447	502	439
5	Riau	1.216	1.256	1.219	972	932	1.028	1.010	968	1.017	1.004	1.120	1.054
6	Kepri	182	214	223	227	227	219	211	209	278	239	212	218
7	Jambi	324	324	447	407	362	386	343	350	365	340	342	378
8	Sumsel	944	959	1.065	1.071	944	1.117	863	897	1.152	1.017	994	1.104
9	Bengkulu	281	254	265	326	220	244	174	194	217	273	256	239
10	Kep Bangka Belitung	131	146	175	143	104	158	113	124	136	111	130	107
11	Lampung	814	829	1.014	854	764	831	570	699	709	783	783	738
12	Banten	357	361	421	442	319	386	280	365	376	405	377	323
13	Metro Jaya	1.836	1.681	2.013	1.819	1.811	1.980	1.870	2.313	2.177	3.062	2.565	2.517
14	Jabar	1.599	1.649	2.201	2.076	1.628	1.781	1.366	2.093	1.935	1.965	2.476	2.315
15	Jateng	907	863	1.053	931	704	702	611	685	732	687	693	622
16	Jatim	2.505	2.408	2.930	2.556	2.297	2.604	1.700	2.198	2.840	2.713	2.373	2.632
17	DIY	488	496	488	494	392	475	314	414	452	396	427	439
18	Bali	276	264	287	271	302	277	227	224	251	242	179	185
19	NTT	569	449	562	580	401	446	411	474	312	335	289	347
20	NTB	398	313	563	355	642	517	423	406	439	488	716	434
21	Kalbar	270	287	351	345	310	328	254	269	333	327	241	297
22	Kaltara	111	90	90	80	76	79	55	82	88	108	96	66
23	Kalteng	239	187	281	250	259	191	173	173	209	248	182	230
24	Kalsel	339	420	398	446	354	382	355	331	413	404	392	304
25	Kalim	333	328	301	303	264	305	211	302	371	301	230	218
26	Sulsel	1.102	1.715	1.447	1.238	1.343	1.376	1.152	1.109	1.217	1.662	1.731	1.840
27	Sulbar	104	147	139	166	144	138	97	148	142	137	131	133
28	Sulut	673	570	633	681	714	821	698	629	682	916	760	852
29	Sulteng	476	409	459	421	356	405	349	386	419	298	320	340
30	Sultra	308	335	256	449	242	343	267	223	207	265	251	240
31	Gorontalo	261	208	232	231	219	219	197	170	203	194	190	204
32	Maluku	227	214	217	171	159	173	145	183	206	238	203	151
33	Malut	98	96	93	102	116	125	107	92	61	70	75	88
34	Papua	394	290	394	299	163	237	281	327	421	346	364	380
35	Papua Barat	195	148	148	302	336	260	143	223	217	179	123	179
	JUMLAH	22.226	22.456	25.114	23.731	21.303	23.151	19.258	21.497	23.215	24.665	24.300	24.072

Gambar 1. Data Jumlah Tindak Pidana Menurut Kepolisian Daerah[2]

No	POLDA	PNS	TNI Polri	Pelajar Mahasiswa	Tani Nelayan Pedagang	Guru Dosen	Tokoh Agama	Karyawan Swasta	Buruh Supir	Wira swasta	Lain nya	Tidak Bekerja
(1)	(2)	(3)										
1	Bareskrim Polri	5	2	19	8	3	0	69	11	185	732	21
2	Polda Aceh	197	66	718	1.376	25	0	2.496	216	30	3.907	293
3	Polda Sumut	430	178	1.108	3.287	91	0	9.214	1.366	52	24.727	1.088
4	Polda Sumbar	64	50	403	785	20	0	1.625	428	16	3.288	371
5	Polda Riau	50	30	251	813	19	0	2.283	621	23	7.009	228
6	Polda Kep Riau	13	6	24	51	2	0	190	88	17	1.776	31
7	Polda Jambi	49	19	181	589	3	0	1.079	270	17	2.345	127
8	Polda Sumsel	128	84	318	1.457	31	0	2.083	1.174	13	7.637	362
9	Polda Bengkulu	37	21	97	286	0	0	544	57	3	1.684	32
10	Polda KB Belitung	10	13	61	69	1	0	186	312	3	717	35
11	Polda Lampung	163	47	319	693	30	0	1.977	672	8	4.889	175
12	Polda Banten	31	13	160	70	7	0	863	299	9	3.060	83
13	Polda Metro Jaya	11	19	264	78	8	0	1.453	509	13	23.451	92
14	Polda Jawa Barat	107	32	498	263	25	0	3.612	1.305	87	15.091	309
15	Polda Jateng	89	57	552	428	24	0	3.740	952	19	3.623	134
16	Polda Jatim	209	94	875	1.463	76	0	10.567	516	45	16.025	448
17	Polda DIY	33	27	291	101	16	0	1.312	271	9	2.765	79
18	Polda Bali	22	6	198	175	6	0	1.228	135	5	1.443	57
19	Polda NTB	61	33	202	631	22	0	1.183	178	40	2.373	135
20	Polda NTT	160	47	253	1.317	40	0	909	134	2	3.552	138
21	Polda Kalbar	33	11	170	315	3	0	1.170	92	17	1.627	94
22	Polda Kaltara	10	4	21	75	0	0	185	33	2	709	16
23	Polda Kalteng	40	5	106	231	7	0	1.139	138	1	1.186	87
24	Polda Kalsel	37	20	147	274	3	0	1.748	285	9	2.397	121
25	Polda Kaltim	20	16	130	219	1	0	1.534	158	23	1.815	131
26	Polda Sulsel	259	99	623	1.311	52	0	3.386	736	41	10.523	823
27	Polda Sulbar	20	16	72	237	4	0	327	59	1	1.242	37
28	Polda Sulut	141	47	225	818	31	0	1.249	318	13	8.773	385
29	Polda Sulteng	55	32	64	355	7	0	499	84	12	3.126	76
30	Polda Sultra	142	32	164	260	6	0	867	94	7	2.223	95

No	POLDA	PNS	TNI Polri	Pelajar Mahasiswa	Tani Nelayan Pedagang	Guru Dosen	Tokoh Agama	Karyawan Swasta	Buruh Supir	Wira swasta	Lain nya	Tidak Bekerja
(1)	(2)	(3)										
31	Polda Gorontalo	80	35	84	320	17	0	400	80	2	2.426	97
32	Polda Maluku	77	56	74	173	8	0	228	48	4	1.984	44
33	Polda Maluku	45	29	93	113	8	0	299	25	3	995	34
34	Polda Papua	104	58	77	65	6	0	479	16	1	2.753	56
35	Polda Papua Barat	36	17	40	31	1	0	194	19	1	2.257	12
JUMLAH		2.968	1.321	8.882	18.737	603	0	0	11.699	733	174.130	6.346

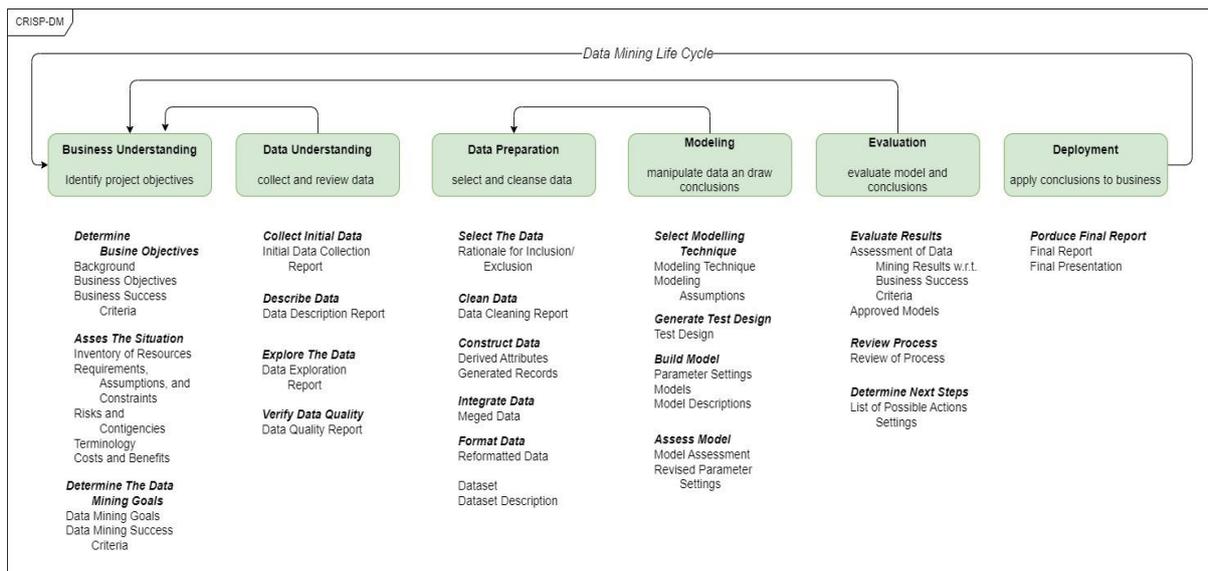
Gambar 2. Data Tingkat Pidana Kejahatan Berdasarkan Pekerjaan Korban Per Polda [2]

Dimana untuk atribut pada jumlah tindak pidana kejahatan data yang digunakan adalah tahun 2021.

2. METODE PENELITIAN

Untuk menangani permasalahan dalam tinjauan ini, digunakan pendekatan teknik CRISP-DM, CRISP-DM adalah sebuah kerangka kerja standar dalam bidang *data mining* yang pertama kali dikonseptualisasikan

oleh tiga inisiatif utama dalam industri *data mining*: Daimler Chrysler (Daimler-Benz), SPSS (ISL), dan NCR. Pendekatan ini terdiri dari enam tahap, yaitu *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment*[13]. Siklus sistematis ini terdiri dari 6 fase yang dapat dijelaskan lebih lanjut dalam Gambar 3.



Gambar 3. CRISP-DM [14]

2.1 Business Understanding

Pada tahap interpretasi bisnis dalam penelitian ini, dilakukan klastering terhadap tindak pidana yang terjadi di Indonesia berdasarkan tahun kejadian dan status pekerjaan pelaku, khususnya jika mereka tidak bekerja. Oleh karena itu, dalam tahap ini penting untuk melakukan penafsiran terkait dengan tujuan, konteks, dan hasil yang ditemukan dalam penelitian ini dalam kerangka pemahaman bisnis. Bagian-bagian yang terkandung dalam pemahaman bisnis ini melibatkan:

2.1.1 Determine Business Objectives

Dengan melakukan klastering dengan data yang dihasilkan oleh mengenai tindak pidana, tidak bekerja pada pelaku diharapkan menghasilkan identifikasi pola tindak pidana, penentuan prioritas tindak pidana, evaluasi kebijakan keamanan, perencanaan strategis.

2.1.2 Asses The Situation

Untuk melihat trend tindak pidana yang terjadi dan jumlah sebarannya yang ada di wilayah kepolisian republik Indonesia maka data yang di gunakan pada tahun 2021 dan pekerjaan pelaku yaitu tidak bekerja.

2.1.3 Determine The Data Mining Goals

Dengan dilakukan proses *data mining* ini adalah untuk mendapatkan informasi baru dengan salah satu algoritma *klastering* sehingga dapat menghasilkan suatu pola tindak pidana sehingga dapat membantu pihak kepolisian untuk

penentuan tindak pidana, evaluasi dan kebijakan keamanan

2.2 Data Understanding

Fase pemahaman data adalah tahap awal dalam proses data mining yang dimulai dengan pengumpulan data. Selanjutnya, tahap ini melibatkan kegiatan seperti mengidentifikasi sifat data, mengenali masalah yang berkaitan dengan kualitas data, meraih wawasan awal dari data, serta menemukan subset data yang menarik sebagai dasar untuk membentuk hipotesis yang akan menghasilkan informasi penting [15]. Data yang digunakan berasal dari data pusiknas bareskrim polri . Dan pada *data understanding* ini memiliki tahapan yaitu :

2.2.1 Collect The Initial Data

Informasi yang dikumpulkan berasal dari . infromasi tindak pidana dengan rentang waktu pada tahun 2021 dan pekerjaan pelaku yaitu tidak bekerja.

2.2.2 Describe The Data

Untuk mengambil data, download data di jurnal tahunan Pusiknas Bareskrim Polri tahun 2021 lalu mencari data yang di dibutuhkan di mana disini ada 2 data yang diambil yaitu jumlah tindak pidana pada tahun 2021 dan data tindak pidana berdasarkan pekerjaan pelaku yaitu tidak bekerja setelah file di *download* maka data di olah terlebih dahulu di *Microsoft Excel* lalu digunakan data tahun 2021. Lalu *download* data pekerjaan pelaku data yang diambil adalah tidak bekerja lalu dengan *Microsoft*

Excel dan ambil atribut tidak bekerja pada bulan tahun 2021. Setelah data diambil maka data tersebut digabungkan sehingga menjadi *dataset* baru dengan atribut nama provinsi, kasus tahun 2021, tidak bekerja. Total data 34 provinsi yang akan dilakukan klustering.

2.2.3 Explore The Data

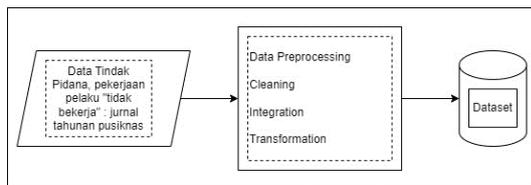
Pada proses *describe the data* ditemukan atribut yang tidak digunakan karena menggunakan presentase sehingga data data dapat dihilangkan.

2.2.4 Verify Data Quality

Pada data ditemukan ketidakcocokan urutan provinsi pada data tindak pidana, tingkat pengangguran sehingga dilakukan pengurutan sesuai dengan 2 data tersebut.

2.3 Data Preparation

Dalam tahap ini, data dipilih dan disaring[16]. Diagram yang digunakan untuk mengilustrasikan tahapan *data preparation* ditunjukkan oleh gambar 4



Gambar 4. Flow Data Preparation [17]

2.3.1 Select Data

Tabel 1 : Atribut Yang Digunakan

No	Atribut	Tipe Atribut	Definisi
1	Nama Satker	<i>Polynomial</i>	Menampilkan nama daerah
2	Jumlah Tindak Pidana	Int	Menampilkan kasus tindak pidana
3	Tidak Bekerja	Int	Menampilkan data pelaku kejahatan berdasarkan pekerjaan

2.3.2 Clean Data

Pada 3 sumber data ini tidak ada data yang kosong dan tidak mengalami proses pembersihan data.

2.3.3 Construct Data

Tabel 2 : Jumlah Tindak Pidana Kejahatan

No	Satker	JAN	...	DES
----	--------	-----	-----	-----

1	Bareskrim Polri	54	60
2	Aceh	633	624
3
35	Papua Barat	195	179

Tabel 3 : Jumlah Tingkat Pelaku Kejahatan Berdasarkan Pekerjaan

No	POLDA	PNS	Tidak Bekerja
1	Bareskrim Polri	5	21
2	Aceh	197	293
3
35	Papua Barat	36	12

2.3.4 Integrate Data

Tahapan ini menggabungkan 2 sumber informasi dalam tabel menjadi satu tabel yang akan ditangani. Hasil informasi terkoordinasi diperkenalkan seperti yang ditampilkan pada Tabel 5.

Tabel 5 : Tabel *integrate data*

No	Satker	Jumlah Tindak Pidana Kejahatan			Jumlah Tingkat Pelaku Berdasarkan pekerjaan		
		JAN	DES	PNS	Tidak Bekerja
1	Bareskrim Polri	54	60	5	21
2	Aceh	633	624	197	293
3
35	Papua Barat	195	179	36	12

2.3.5 Format Data

Format Data dilakukan dengan membuat tabel lain sebagai tipikal jumlah tindak pidana selama tahun 2021 dan digabungkan dengan jumlah tingkat pelaku. ditampilkan pada tabel 6.

Tabel 6 : Tabel format data

No	Kepolisian Daerah	Jumlah tindak pidana kejahatan	Tidak Bekerja
1	Bareskrim Polri	739	21
2	Aceh	8005	293

3
35	Papua Barat	2453	12

2.3.6 Modelling

Tahap ini langsung terkait dengan perancangan strategi data mining. Ini mencakup pemilihan metode data mining, perhitungan, dan pengambilan keputusan terkait dengan batasan kualitas yang optimal [18]. Tahapan dalam tampilan pemodelan adalah sebagai berikut.

2.3.7 Select Modelling Technique

Prosedur data mining yang dipilih adalah klustering dimana perhitungan yang digunakan adalah K-Medoids dan K-Means.

2.3.8 Generate Test Design

Metode klustering tidak melibatkan tahap pelatihan karena klustering merupakan bentuk *unsupervised learning*. Ini melakukan pengelompokan berdasarkan kemiripan atribut tanpa memerlukan pembelajaran sebelumnya, yang merupakan karakteristik uniknya dalam konteks strategi klustering lainnya.

2.3.9 Build Model

Tahap ini berhasil dijalankan dengan bantuan aplikasi Rapid Miner, di mana data diproses menggunakan dua teknik perhitungan, yaitu K-Medoids dan K-Means.

2.3.10 Assess Model

Evaluasi kluster yang optimal dilakukan dengan membandingkan dua perhitungan, yaitu K-Means dan K-Medoids, dan hasil kluster terbaik akan dipilih untuk digunakan.

2.4 Evaluation

Tahap interpretasi hasil *data mining* berlangsung pada titik ini. Untuk memastikan bahwa hasil dari fase pemodelan sesuai dengan tujuan fase *business understanding*, evaluasi mendalam dilakukan. Latihan yang diselesaikan pada tahap ini adalah sebagai berikut:

2.4.1 Evaluate Results

Pada tahap ini, efektivitas hasil pemodelan data mining dalam mencapai tujuan yang telah ditetapkan dalam tahap pemahaman bisnis dinilai.

2.4.2 Review Process

Pada tahap ini, proses secara menyeluruh dianalisis kembali dari awal untuk memastikan bahwa tidak ada elemen penting yang terlewat atau diabaikan dalam proses ini.

2.4.3 Determine Next Steps

Pada tahap ini, dilakukan penentuan terkait langkah selanjutnya, yang terdiri dari dua opsi saja: kembali ke tahap awal (Bisnis Pemahaman) atau melanjutkan ke tahap akhir (Implementasi/Deployment).

2.5 Deployment

Pada tahap *deployment*, tindakan yang dilakukan adalah menghasilkan laporan yang menjelaskan implementasi proses *data mining* dan menentukan langkah-langkah selanjutnya yang perlu diambil[19]. Laporan ini akan digunakan sebagai dasar untuk merumuskan prosedur pencegahan dan penanganan tindak kriminal.

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Tahap ini mengilustrasikan tujuan utama penelitian, yakni untuk membandingkan pengelompokan (klustering) tindak pidana yang terjadi di seluruh wilayah hukum Kepolisian Republik Indonesia. Data yang digunakan berasal dari Jurnal Tahunan Pusiknas Bareskrim POLRI tahun 2021. Hasil dari pengelompokan ini akan memberikan wawasan yang berguna bagi kepolisian dalam merancang strategi penanganan tindak pidana dengan penempatan personil yang tepat di seluruh wilayah hukum Indonesia.

3.2 Data Understanding

Pemahaman data yang dihasilkan pada tahapan ini adalah jumlah tindak pidana kejahatan pada tahun 2021 dan pekerjaan pelaku.

3.3 Data Preparation

Informasi yang diperoleh dari tahap persiapan data adalah informasi terakhir yang akan dimanfaatkan dalam proses *data mining*. Dampak lanjutan dari tahap ini bahkan tercermin dalam struktur, seperti yang terlihat dalam Tabel 7.

Tabel 7 : Tabel *Data Preparation*

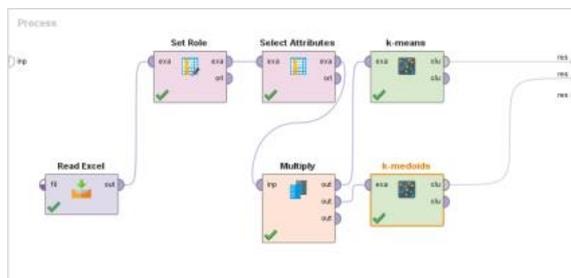
No	Kepolisian Daerah	Jumlah tindak pidana kejahatan	Tidak Bekerja
1	Bareskrim Polri	739	21
2	Aceh	8005	293
3
35	Papua Barat	2453	12

3.4 Modelling

Pada tahap ini, penggunaan aplikasi Rapid Miner secara langsung terlibat dalam mendukung proses *data mining*. Tahap pemodelan terbagi menjadi empat fase, di mana dua fase sudah dieksplorasi, terutama yang terfokus pada fase pertama dan selanjutnya fase kedua. Fokus dua fase yang berbeda yang akan dibahas adalah tahap pembuatan model dan evaluasi model.

3.4.1 Build Model

Aplikasi Rapid Miner digunakan untuk mengolah data tindak pidana yang akan mengalami proses klustering. Algoritma yang diterapkan adalah K-Means dan K-Medoids. Rancangan model ini terlihat dalam Gambar 5.



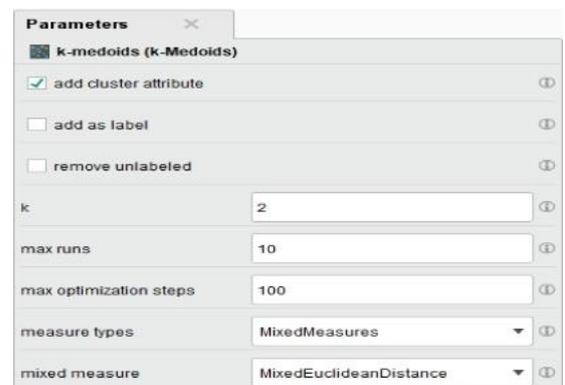
Gambar 5. Build Model

Proses *build model* menggunakan Rapid Miner, Rapid Miner adalah perangkat lunak yang dirancang oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Ralf Klinkenberg dari rapid-i.com. Perangkat lunak ini dilengkapi dengan antarmuka pengguna berbasis grafis (*Graphical User Interface* atau GUI), yang bertujuan untuk memberikan kemudahan bagi pengguna dalam menggunakan aplikasi ini [20]. Yang pertama dilakukan adalah menggunakan operator read excel yang berfungsi untuk mengolah data excel tindak pidana. Lalu data akan dilakukan *set role* setelah dilakukan *set role* dipilih atribut yang akan digunakan menggunakan operator *select attributes*, data yang sudah dipilih atribut nya akan diolah oleh operator *multiply* yang berfungsi untuk memproses

algoritma lebih dari satu yaitu *K-Means* dan *K-Medoids*. Parameter yang di gunakan oleh *K-Means* terdapat pada gambar 7 dan *K-Medoids* terdapat pada gambar 6



Gambar 6. Parameter *K-Means*



Gambar 7. Parameter *K-Medoids*

3.4.2 Asses Model

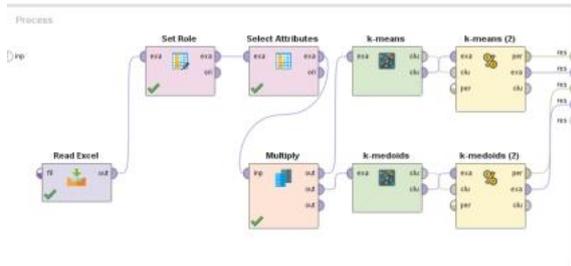
Pembuatan model dilakukan dengan variasi dari 2 hingga 8 kelompok kluster untuk mencari nilai terbaik menggunakan *indeks Davies-Boulding*.

3.5 Evaluation

Setiap kluster yang terbentuk akan dilakukan analisis dan evaluasi dengan tujuan untuk mencari nilai kluster terbaik

3.5.1 Evaluate Results

Indeks Davies-Bouldin menghasilkan nilai numerik, dan nilai yang lebih kecil mengindikasikan kualitas pengelompokan yang lebih baik. Dalam penelitian ini, untuk menentukan nilai terbaik, dilakukan pengujian dengan jumlah kluster mulai dari 2 hingga 8, dan kemudian memilih hasil *indeks Davies-Bouldin* yang paling kecil, seperti yang ditunjukkan pada Gambar 8.



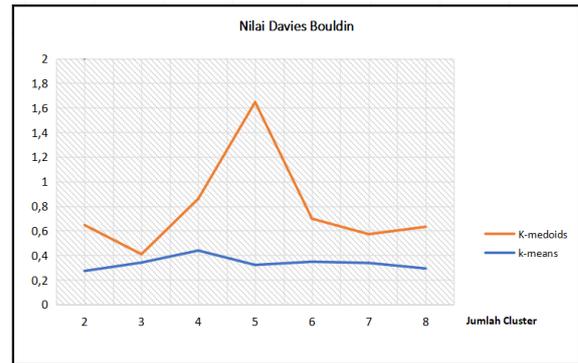
Gambar 8. Testing Performace

Hasil dari proses menggunakan Rapid Miner menghasilkan performa *data mining* dengan algoritma K-Means dan K-Medoids yang tersedia dalam Tabel 8.

Tabel 8 : Tabel *Data Performance* DBI

No	klaster	K-Medoids	K-Means
1	2	0,645	0,272
2	3	0,408	0,34
3	4	0,859	0,438
4	5	1,645	0,321
5	6	0,697	0,347
6	7	0,571	0,337
7	8	0,631	0,292

Performance data mining yang sudah dalam bentuk table dibuat *visualisasi* nya menggunakan *chart* agar memudahkan dalam melakukan analisis yang di tunjukan pada gambar 9.



Gambar 9. Grafik nilai DBI K-Means dan K-Medoids

Dari grafik yang di dihasilkan dapat dilihat bahwa klaster terbaik adalah menggunakan algoritma K-medoids dengan jumlah klaster 2 dan nilai Davies Bouldin Indeks adalah 0,645 sedangkan pada algoritma *K-Means* dengan jumlah klaster 2 mendapatkan nilai 0,272. Reliabilitas klasifikasi indeks atau bisa disebut Davies Bouldin Indeks (DBI) dengan menentukan indeks terkecil [18], [21]. Dengan hasil tersebut nilai yang mendekati 0 adalah algoritma *K-Means* maka model dengan klaster 2 ditunjukan pada gambar 10

Cluster Model

```
Cluster 0: 5 items
Cluster 1: 30 items
Total number of items: 35
```

Gambar 10. klaster K-Means dengan nilai DBI terkecil

Setelah diketahui nilai klaster yang terbaik dilakukan analisis dengan jumlah 2 klaster yang terbentuk. Hasil analisis terdapat pada tabel 9

Tabel 9 : Tabel klaster yang Terbetuk

Klaster	Satker	Jumlah Tindak Pidana	Tidak Bekerja
klaster-1	Sumut	39816,0	1088,0
klaster-1	Metro Jaya	25644,0	92,0
klaster-1	Jabar	23084,0	309,0
klaster-1	Jatim	29756,0	448,0
klaster-1	Sulsel	16932,0	823,0
klaster-2	Bareskrim Polri	739,0	21,0
klaster-2	Aceh	8005,0	293,0
klaster-2	Sumbar	6352,0	371,0
klaster-2	Riau	12796,0	228,0
klaster-2	Kepri	2659,0	31,0
klaster-2	Jambi	4358,0	127,0
klaster-2	Sumsel	12127,0	362,0
klaster-2	Bengkulu	2943,0	32,0

klaster-2	Kep Bangka Belitung	1578,0	35,0
klaster-2	Lampung	9388,0	175,0
klaster-2	Banten	4412,0	83,0
klaster-2	Jateng	9390,0	134,0
klaster-2	DIY	5275,0	79,0
klaster-2	Bali	2985,0	57,0
klaster-2	NTT	5175,0	138,0
klaster-2	NTB	5694,0	135,0
klaster-2	Kalbar	3612,0	94,0
klaster-2	Kaltara	1021,0	16,0
klaster-2	Kalteng	2622,0	87,0
klaster-2	Kalsel	4538,0	121,0
klaster-2	Kaltim	3523,0	131,0
klaster-2	Sulbar	1626,0	37,0
klaster-2	Sulut	8629,0	385,0
klaster-2	Sulteng	4638,0	76,0
klaster-2	Sultra	3384,0	95,0
klaster-2	Gorontalo	2528,0	97,0
klaster-2	Maluku	2287,0	44,0
klaster-2	Malut	1123,0	34,0
klaster-2	Papua	3896,0	56,0
klaster-2	Papua Barat	2453,0	12,0

Dari hasil tabel 9 dapat di ketahui bahwa hasil yang didapat 2 klaster dengan masing-masing klaster 1 dengan 5 anggota dan klaster 2 dengan 30 anggota maka dari itu dilakukan pemecahan klaster besar pada klaster 0 dengan dilakukan proses klastering menjadi 2 klaster

3.5.2 Review Process

Seluruh tahap siklus telah diselesaikan sesuai dengan metode yang digunakan, yaitu CRISP-DM, sehingga kita dapat melanjutkan ke tahap berikutnya.

3.5.3 Determine Next Steps

Tahap ini akan menentukan apakah kita akan melanjutkan ke tahap berikutnya atau memeriksa kembali beberapa hal. Semua tahapan siklus telah diselesaikan dan telah diberikan informasi serta wawasan baru mengenai pengelompokan pelanggaran yang terjadi di wilayah Indonesia. Oleh karena itu, kita dapat memutuskan untuk melanjutkan ke tahap berikutnya, terutama tahap Implementasi (*Deployment*).

3.5.4 Deployment

Tahap ini merupakan proses pembuatan laporan atau, dengan kata lain, penerapan hasil dari Data Mining. Hasil yang ditemukan dalam pengelompokan data tindak pidana menunjukkan

bahwa klaster 2 adalah klaster yang memiliki tingkat tindak pidana tertinggi. Wilayah ini memiliki total kasus sebanyak 139,756, dengan total pelaku yang tidak bekerja sebanyak 3,856. Wilayah-wilayah ini mencakup Bareskrim, Aceh, Sumbar, Riau, Kepri, Jambi, Sumsel, Bengkulu, Kepulauan Bangka Belitung, Lampung, Banten, Jawa Tengah, Daerah Istimewa Yogyakarta, Bali, Nusa Tenggara Timur, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Utara, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Barat, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua, dan Papua Barat. Strategi yang direkomendasikan adalah menempatkan lebih banyak personel kepolisian di wilayah-wilayah ini yang memiliki tingkat tindak pidana tinggi.

4. SIMPULAN DAN SARAN

Hasil penelitian ini didapatkan hasil perbandingan *klastering* menggunakan algoritma *K-Means* dan *K-Medoids* dengan menggunakan Teknik *data mining CRISP-DM*. Dengan menggunakan *dataset* yang terdiri dari 2 atribut yaitu jumlah tindak pidana, tidak bekerja. Dari hasil penelitian tersebut dengan menggunakan pengujian Davies Bouldin Indeks dengan menggunakan algoritma *K-Medoids* dan *K-Means* didapatkan hasil terbaik pada 2 *klaster*, algoritma *K-Medoids* dengan nilai indeks sebesar 0,645 dan *K-Means* 0,272 sehingga algoritma *K-*

Means menjadi nilai terbaik pada Davies Bouldin Indeks. klastering dengan data tersebut menghasilkan data tindak pidana sebanyak 2 klaster dengan tindak pidana tinggi pada klaster 2 dengan total 139756 dan tidak bekerja 3586 kasus terdapat pada 30 provinsi yaitu Bareskrim, Aceh, Sumbar, Riau, Kepri, Jambi, Sumsel, Bengkulu, Kep Bangka Belitung, Lampung, Banten, Jateng, DIY, Bali, NTT, NTB, Kalbar, Kaltara, Kalteng, Kalsel, Kaltim, Sulbar, Sulut, Sulteng, Sultra, Gorontalo, Maluku, Malut, Papua, Papua Barat. Dan 5 provinsi dengan total kasus 135232 dan tidak bekerja 2760

orang, kasus nilai tindak pidana rendah yaitu Sumut, Metro Jaya, Jabar, Jatim, Sulsel. Dengan data tersebut diharapkan kepolisian dapat menerapkan kebijakan untuk menempatkan personilnya di wilayah-wilayah yang rawan dengan tindak pidana agar masyarakat dapat semakin merasa aman di wilayah hukum masing-masing di provinsi Indonesia. Saran dalam penelitian ini yaitu dalam membandingkan model algoritma *K-Means* dan *K-Medoids* dapat di tambahkan atribut lain untuk melihat hasil yang optimal pada nilai DBI dan pembentukan klaster yang lebih baik.

DAFTAR PUSTAKA

- [1] D. Dwirohayati, N. Herawadi Sudibyo, and A. Kejahatan, "Sistem Informasi Pemetaan Wilayah Rawan Kriminalitas Polresta Bandar Lampung Menggunakan K-Means Clustering," *IJCCS*, vol. x, No.x, no. 0721, p. 1, 2020.
- [2] Pusiknas Bareskrim Polri, "Jurnal Tahunan Pusiknas Bareskrim Polri," *Jurnal Data Pusat Informasi Kriminal Nasional Bareskrim Polri*, p. 155, 2021, [Online]. Available: https://pusiknas.polri.go.id/web_pusiknas/laporan/1649645185073.pdf
- [3] M. A. Rizaty, "Berapa jumlah Polisi di Indonesia?" Accessed: Jul. 17, 2023. [Online]. Available: <https://dataindonesia.id/varia/detail/berapa-jumlah-polisi-di-indonesia>
- [4] F. Nurdiyansyah and I. Akbar, "Implementasi Algoritma K-Means untuk Menentukan Persediaan Barang pada Poultry Shop," *Jurnal Teknologi dan Manajemen Informatika*, vol. 7, no. 2, pp. 86–94, 2021, doi: 10.26905/jtmi.v7i2.6377.
- [5] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus: Poli Klinik PT.Inecda)," *Jurnal Ilmu Komputer*, vol. 9, no. 2, pp. 148–153, Oct. 2020, doi: 10.33060/jik/2020/vol9.iss2.181.
- [6] W. Saputro, M. Reza Pahlevi, and A. Wibowo, "Analisis Algoritma K-Means Untuk Klasterisasi Tindak Pidana Korupsi Di Wilayah Hukum Indonesia," *JIKO (Jurnal Informatika dan Komputer)*, vol. 3, no. 3, pp. 137–142, 2020, doi: 10.33387/jiko.v3i3.1960.
- [7] J. J. Purnama, "Analisa Algoritma K-Means Clustering Pemetaan Jumlah Tindak Pidana," *KLIK - KUMPULAN JURNAL ILMU KOMPUTER*, vol. 6, no. 2, p. 128, 2019, doi: 10.20527/klik.v6i2.208.
- [8] U. T. Suryadi and Y. Supriatna, "Sistem Clustering Tindak Kejahatan Pencurian Di Wilayah Jawa Barat Menggunakan Algoritma K-Means," *JTIK*, vol. 22, no. April, p. 111, Apr. 2019.
- [9] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. I. R.H Zer, and D. Hartama, "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia," *Jurnal Teknologi Informasi*, vol. 4, no. 1, pp. 128–132, 2020, doi: 10.36294/jurti.v4i1.1266.
- [10] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, and D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *Jurnal Teknologi Informasi*, vol. 4, no. 1, pp. 166–173, 2020, doi: 10.36294/jurti.v4i1.1296.
- [11] Z. Wahidah and D. T. Utari, "Implementation K-Means Algorithm to Group Provinces By Factors Influenced Criminal Act in Indonesia in 2019," *ENTHUSIASTIC*, vol. 2, no. 1, pp. 37–46, 2022, [Online]. Available: <https://journal.uui.ac.id/ENTHUSIASTIC>

- [12] N. Arief, I. Sudahri Damanik, E. Irawan, S. Tunas Bangsa, and S. Utara, "KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Algoritma K-Medoids Dalam Mengelompokkan Tingkat Kasus Kejahatan di Setiap Provinsi," *Media Online*, vol. 2, no. 3, pp. 111–116, 2021, [Online]. Available: <https://djournals.com/klik>
- [13] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [14] J. Rochel and F. Evéquo, "Getting into the engine room: a blueprint to investigate the shadowy steps of AI ethics," *AI Soc*, vol. 36, no. 2, pp. 609–622, Jun. 2021, doi: 10.1007/s00146-020-01069-w.
- [15] D. Kurniawan and M. Yasir, "Optimization Sentimen Analysis using CRISP-DM and Naive Bayes Methods Implemented on Social Media," *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, vol. 6, no. 2, p. 74, 2022, doi: 10.22373/cj.v6i2.12793.
- [16] S. Shedriko and M. Firdaus, "Penentuan Klasifikasi Dengan CRISP-DM Dalam Memprediksi Kelulusan Mahasiswa Pada Suatu Mata Kuliah," *Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi)*, vol. 6, no. 1, 2022, doi: 10.30998/semnasristek.v6i1.5814.
- [17] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "Data Mining Clustering Food Expenditure In Indonesia," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 6, pp. 1491–1500, Dec. 2022, doi: 10.20884/1.jutif.2022.3.6.331.
- [18] Y. Sopyan, A. D. Lesmana, and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2697.
- [19] S. Navisa, Luqman Hakim, and Aulia Nabilah, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM," *Jurnal Sistem Cerdas*, vol. 4, no. 2, pp. 114–125, 2021, doi: 10.37396/jsc.v4i2.162.
- [20] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.
- [21] I. W. Septiani, Abd. C. Fauzan, and M. M. Huda, "Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 3, no. 4, p. 556, Jul. 2022, doi: 10.30865/json.v3i4.4055.