



Pemetaan Daerah Produksi Perkebunan Kelapa Sawit Pada Provinsi Indonesia Menggunakan Algoritma *K-Medoids*

Fatma Eka Zulfiakhoir¹, Yuyun Umaidah², Purwantoro³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

Abstract

Received: 11 Agustus 2022
Revised: 14 Agustus 2022
Accepted: 19 Agustus 2022

In the National Leading Plantation Statistics yearbook issued by the Directorate General of Plantation, Ministry of Agriculture of the Republic of Indonesia, it is stated that Indonesia is the world's number one producer of palm oil and also the owner of the largest oil palm plantation area in the world. Indonesia's palm oil production reached 43.5 million tons, with an average growth percentage of 3.61%. Therefore, it is necessary to map oil palm plantations from each province of Indonesia so that it can be seen which areas have the potential to produce oil palm. Thus, later the area can be maximized again productivity. Then, to find out also which provinces have a low potential level of palm oil production so that later the productivity of the area can be assisted by analyzing what patterns can be found from the highest potential level of oil palm production. In this research, mapping will be carried out using clustering techniques in data mining, using the k-medoid algorithm. The results showed that there were 3 clusters, namely cluster 1 (category of provinces with low production areas of oil palm), cluster 2 (category of provinces with high production areas of oil palm), and cluster 3 (category of provinces with medium production areas of oil palm).). The silhouette coefficient evaluation result in the model calculation using this technique is 0.64. This range value is included in the medium structure criteria (good cluster structure).

Keywords: Oil Palm, Data Mining, Clustering, K-Medoid

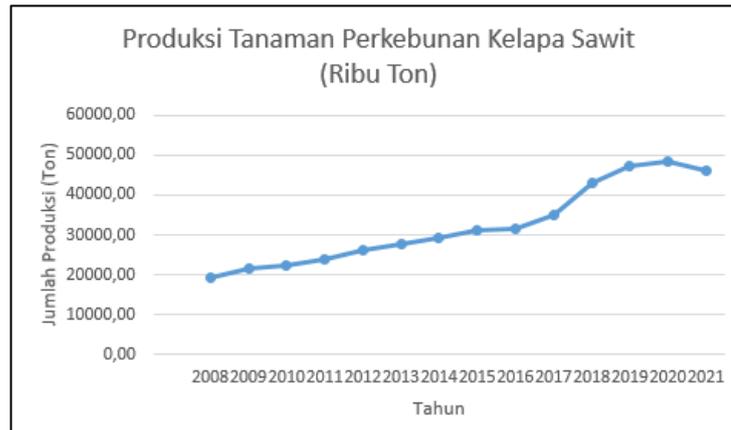
(*) Corresponding Author: fatma.eka18170@student.unsika.ac.id, yuyun.umaidah@staff.unsika.ac.id, purwantoro.masbro@staff.unsika.ac.id

How to Cite: Zulfiakhoir, F., Umaidah, Y., & Purwantoro, P. (2022). Pemetaan Daerah Produksi Perkebunan Kelapa Sawit Pada Provinsi Indonesia Menggunakan Algoritma K-medoids. *Jurnal Ilmiah Wahana Pendidikan*, 8(16), 195-208. <https://doi.org/10.5281/zenodo.7067527>

PENDAHULUAN

Dalam buku tahunan Statistik Perkebunan Unggulan Nasional 2019-2021 yang dikeluarkan oleh Direktorat Jenderal Perkebunan Kementerian Pertanian Republik Indonesia menyebutkan bahwa Indonesia sebagai negara peringkat satu produsen kelapa sawit dunia dan juga pemilik luas areal perkebunan kelapa sawit terbesar di dunia.

Menurut berita harian kompas.com pada laman *website*-nya menulis bahwa produksi sawit Indonesia dalam tahun mencapai 43,5 juta ton, dengan persentase pertumbuhan rata-rata sebesar 3,61%. Sehingga, produksi sawit yang merupakan bahan dasar utama dari minyak sawit pun ikut serta meroket. Berikut ini Gambar 1 yang menunjukkan grafik mengenai peningkatan produksi kelapa sawit tiap tahunnya dalam rentang waktu 2008 s.d 2021.



Gambar 1. Produksi Tanaman Perkebunan Kelapa Sawit Tahun 2008-2021
(Sumber: Badan Pusat Statistik, 2021)

Dapat dilihat dari gambar grafik diatas bahwa adanya peningkatan produksi tanaman perkebunan kelapa sawit di Indonesia setiap tahunnya sejak tahun 2008 hingga tahun 2020. Namun, dalam waktu dekat ini yaitu pada tahun 2021 terlihat ada penurunan produksi kelapa sawit. Berdasarkan data dari Badan Pusat Statistik (BPS) jumlah produksi kelapa sawit pada tahun 2020 sebanyak 48296,90 (ribu ton), sedangkan untuk tahun 2021 sebanyak 46223,30 (ribu ton). Lebih tepatnya penurunan jumlah produksi kelapa sawit di tahun 2021 yaitu sekitar 4,29%.

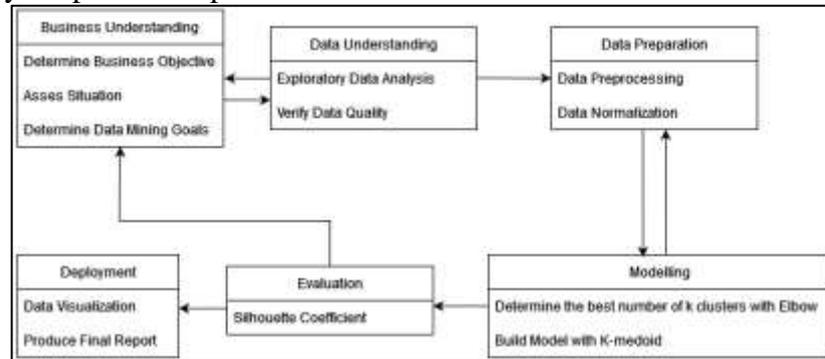
Oleh karena itu, diperlukannya pemetaan perkebunan kelapa sawit dari tiap provinsi Indonesia supaya dapat diketahui daerah mana saja yang potensial untuk memproduksi kelapa sawit. Sehingga, nantinya daerah tersebut dapat dimaksimalkan kembali produktivitasnya. Kemudian, untuk mengetahui juga daerah provinsi mana yang tingkat potensial produksi kelapa sawitnya rendah sehingga nantinya dapat dibantu produktivitas daerah tersebut dengan melakukan analisis pola apa saja yang dapat diketahui dari tingkat potensial produksi kelapa sawit yang paling tinggi.

Pengolahan data pada penelitian ini akan dilakukan menggunakan teknik *clustering* pada *data mining*, dengan menggunakan algoritma *k-medoid*. *Data mining* merupakan teknik yang dapat digunakan untuk menemukan informasi penting dari sejumlah data berskala besar (Bulolo, 2020) .

Berdasarkan penjelasan diatas maka penelitian ini akan melakukan pemetaan terhadap daerah produksi perkebunan kelapa sawit pada provinsi di Indonesia menggunakan algoritma *k-medoids* dan metode evaluasi *silhouette coefficient* dengan *tools Rstudio* dan bahasa pemrograman *R*. Adapun data yang telah dikumpulkan untuk penelitian ini berasal dari laman *website* Badan Pusat Statistik Indonesia yaitu *bps.go.id* dan *website* Kementerian Pertanian Republik Indonesia yaitu *pertanian.go.id* dengan rentang waktu tahun 2008 s.d 2021. Hasil pemetaan pada penelitian ini nantinya akan di visualisasikan menggunakan *software quantum geographic information system (QGIS)*.

METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan untuk pengolahan data adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Adapun alur tahapan metodenya dapat dilihat pada Gambar 2 berikut ini.



Gambar 2. Alur Proses Metode CRISP-DM

Business Understanding

Pada tahap *business understanding* ini ada empat proses yang dilakukan, yaitu:

1. *Determine Business Objective*
Proses yang dilakukan ditahap ini yaitu memahami tujuan yang ingin dicapai dengan melakukan *research* mengenai permasalahan yang ada mengenai kelapa sawit di Indonesia.
2. *Asses Situation*
Proses yang dilakukan di tahap ini mencari di berbagai sumber mengenai situasi dari permasalahan yang ada mengenai kelapa sawit di Indonesia saat ini, seperti dari artikel *journal*, berita *online*, dan sumber lainnya.
3. *Determine Data Mining Goals*
Selanjutnya setelah mengetahui situasi dari permasalahan yang ada saat ini, maka diperlukan penentuan tujuan *data mining* yang akan dilakukan.

Data Understanding

Pada tahap *data understanding* pada penelitian ini ada dua proses yang dilakukan, yaitu:

1. *Exploratory Data Analys (EDA)*
Pada tahap ini akan dilakukan analisis mengenai dataset yang akan digunakan pada penelitian ini. Kemudian, mengidentifikasi *outliers*, *missing values* atau *human error* yang ada dalam data, Intinya memaksimalkan wawasan mengenai dataset sebaik mungkin.
2. *Verify Data Quality*
Pada tahap ini dilakukannya pemeriksaan ulang mengenai kualitas dataset yang akan digunakan.

Data Preparation

Pada tahap ini ada dua langkah yang akan dilakukan pada penelitian ini, yaitu:

1. *Data Preprocessing*
Pada tahap ini melakukan perbaikan masalah dalam data yang sudah teridentifikasi di tahap *data understanding* sebelumnya, seperti mengisi atau membersihkan *missing value*, dan membersihkan data akibat *human error*.
2. *Data Normalization*

Pada proses ini dilakukan kesetaraan pada nilai *variabel* supaya memiliki nilai rentang yang sama, tidak ada yang memiliki nilai terlalu besar maupun memiliki nilai terlalu kecil. Metode normalisasi yang akan digunakan dalam penelitian ini adalah metode *Min-Max Normalization*.

Modelling

Pada tahap *modelling* ini dilakukan penerapan teknik data mining yang digunakan dan penentuan pembuatan model. Sebelum dilakukan pembuatan model langkah awal adalah memilih jumlah *cluster* terbaik dengan menggunakan teknik *elbow*.

Evaluation

Pada tahap ini merupakan tahap evaluasi terhadap hasil pemodelan menggunakan teknik *silhouette coefficient*.

Deployment

Pada tahap ini dilakukan representasi dari pengetahuan yang didapat selama penelitian. Representasi pengetahuan dilakukan menggunakan visualisasi data dari pemetaan produksi kelapa sawit berdasarkan provinsi di Indonesia menggunakan *software* sistem informasi geografis yaitu QGIS.

HASIL DAN PEMBAHASAN

Hasil dari penelitian ini adalah pemetaan daerah produksi kelapa sawit pada provinsi Indonesia menggunakan algoritma *k-medoids clustering*. Kemudian, hasil *clustering* tersebut di evaluasi menggunakan teknik evaluasi model yaitu *silhouette coefficient*. Selanjutnya, hasil *clustering* akan dibuat visualisasi pemetaan menggunakan *software* sistem informasi geografis yaitu QGIS.

Business Understanding

Prospek pertumbuhan produksi kelapa sawit Indonesia menunjukkan angka skala yang terus menaik tiap tahunnya berdasarkan data dari badan pusat statistik. Dalam laman website resmi Gabungan Pengusaha Kelapa Sawit Indonesia (gapki.id) menyebutkan bahwa perkebunan kelapa sawit memiliki kontribusi diberbagai aspek kehidupan seperti aspek ekonomi, yang mana dapat dilihat dalam sumber devisa sawit sebagai penghasil devisa terbesar yang juga meningkatkan pendapatan petani.

Oleh karena itu, penelitian mengenai pemetaan perkebunan kelapa sawit dari tiap provinsi Indonesia sangat penting dilakukan supaya dapat diketahui daerah mana saja yang potensial untuk memproduksi kelapa sawit. Sehingga, nantinya daerah tersebut dapat dimaksimalkan kembali produktivitasnya. Kemudian, untuk mengetahui juga daerah provinsi mana yang tingkat potensial produksi kelapa sawitnya rendah sehingga nantinya dapat dibantu produktivitas daerah tersebut dengan melakukan analisis pola apa saja yang dapat diketahui dari tingkat potensial produksi kelapa sawit yang paling tinggi.

Data Understanding

1. *Exploratory Data Analysis* (EDA)

Data yang dipakai dalam penelitian ini merupakan data produksi tanaman perkebunan dalam rentang waktu tahun 2008 s.d 2021. Data ini diambil dari laman *website* Badan Pusat Statistik Indonesia yaitu *bps.go.id* dan *website* Kementerian Pertanian Republik Indonesia yaitu *pertanian.go.id*. Berikut Gambar 3 yang menunjukkan data dalam penelitian ini.

No.	Provinsi	Produksi Tanaman Perkebunan Kelapa Sawit (Ribu Ton)													
		2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
1	ACEH	344,70	695,60	616,30	592,20	614,80	817,53	845,60	806,50	732,30	887,30	1037,60	1153,30	1134,40	1036,10
2	SUMATERA UTARA	3882,40	3862,40	3899,60	4019,20	3973,42	4340,20	4870,20	5193,10	5683,70	4832,00	5737,50	5647,30	5776,80	5310,50
3	SUMATERA BARAT	981,30	896,30	885,90	827,40	890,12	1022,33	924,80	826,60	1183,30	1209,20	1248,30	1255,40	1112,30	1352,00
4	RIAU	4812,00	5311,40	5496,00	3894,50	6384,54	6647,00	6893,20	8059,80	7668,20	7591,20	8486,00	9512,00	9684,30	8620,10
5	JAMBI	1426,50	1490,90	1044,10	1773,10	1718,29	1749,62	1773,70	1794,80	1435,10	1783,00	2691,30	2886,40	3022,60	2575,10
6	SUMATERA SELATAN	1891,40	2113,50	2542,80	2473,70	2492,90	2690,62	2791,80	2821,80	2929,50	2887,00	1793,60	4049,20	4267,00	3062,40
7	BENGKULU	500,30	736,00	796,00	827,10	802,02	787,05	798,80	747,50	750,28	846,70	1047,70	1032,10	1063,40	1152,70
8	LAMPUNG	416,30	389,30	465,70	424,00	433,82	424,05	455,90	434,30	425,80	450,00	487,20	414,20	384,90	420,70
9	KEP. BANGKA BELITUNG	412,00	448,60	480,20	304,60	304,60	308,13	326,60	323,10	726,60	716,10	902,30	815,70	843,00	800,40
10	KEP. RIAU	10,60	11,30	14,10	15,20	37,20	36,77	45,00	45,10	21,40	25,10	28,90	22,80	20,00	18,00
11	DKI JAKARTA					0,00	0,00	0,00	0,00			900,30			
12	JAWA BARAT	12,60	20,70	16,40	22,10	22,72	32,64	33,00	33,00	32,88	42,40	48,00	32,20	33,00	32,80
13	JAWA TENGAH					0,00	0,00	0,00	0,00						
14	DI YOGYAKARTA					0,00	0,00	0,00	0,00		46,00				
15	JAWA TIMUR					0,00	0,00	0,00	0,00						
16	BANTEN	25,00	25,10	38,00	21,80	27,67	27,08	24,30	25,50	27,50	32,10	38,40	31,30	27,40	30,10
17	BALI					0,00	0,00	0,00	0,00						
18	NUSA TENGGARA BARAT					0,00	0,00	0,00	0,00			38,40			
19	NUSA TENGGARA TIMUR					0,00	0,00	0,00	0,00						
20	KALIMANTAN BARAT	1134,40	1331,70	1426,90	1441,00	1830,80	1794,47	1845,30	2108,10	2192,60	2526,00	3088,90	3255,30	3471,40	3835,90
21	KALIMANTAN TENGAH	1295,70	1798,10	1724,70	2409,30	2938,20	3127,14	3158,20	3573,00	4200,10	5200,90	7230,10	7664,80	7985,80	8600,90
22	KALIMANTAN SELATAN	891,10	1041,40	1049,20	1045,20	1235,08	1244,04	1440,60	1090,30	1730,40	1560,10	1444,20	1605,40	1561,10	1212,80
23	KALIMANTAN TIMUR	338,40	436,40	700,00	893,30	1288,14	1340,78	1487,30	1388,60	2338,40	2318,30	1784,30	3088,00	1821,20	3808,30
24	KALIMANTAN UTARA					104,73	255,20	208,10	167,30	205,20	305,10	281,40	301,80	370,00	

Gambar 3. Data Produksi Tanaman Perkebunan Kelapa Sawit

Dapat dilihat pada Gambar 3 baris data provinsi yang tidak diberi warna merupakan baris data yang siap digunakan pada tahap pemodelan. Baris provinsi yang memiliki warna merah menggambarkan bahwa dalam dataset provinsi tersebut tidak memproduksi tanaman kelapa sawit sepanjang tahun 2008 s.d 2021, yang artinya provinsi tersebut tidak masuk dalam data yang akan dilakukan *clustering*.

Pada baris data provinsi yang berwarna kuning terdapat kejanggalan yang ditemukan. Dapat dilihat bahwa pada data tersebut dalam rentang waktu 2008 s.d 2021 tidak pernah memproduksi tanaman kelapa sawit sama sekali terkecuali pada tahun 2018, yang janggalnya hanya terdapat pada tahun tersebut. Berikut Gambar 4 yang menunjukkan kejanggalan pada data provinsi dengan baris berwarna kuning.

No.	Provinsi	2018
1	ACEH	1037,60
2	SUMATERA UTARA	5737,50
3	SUMATERA BARAT	1248,30
4	RIAU	8486,00
5	JAMBI	2691,30
6	SUMATERA SELATAN	3793,60
7	BENGKULU	1047,70
8	LAMPUNG	487,20
9	KEP. BANGKA BELITUNG	900,30
10	KEP. RIAU	28,90
11	DKI JAKARTA	900,30
12	JAWA BARAT	46,00
13	JAWA TENGAH	
14	DI YOGYAKARTA	46,00
15	JAWA TIMUR	
16	BANTEN	38,40
17	BALI	
18	NUSA TENGGARA BARAT	38,40
19	NUSA TENGGARA TIMUR	
20	KALIMANTAN BARAT	3086,90
21	KALIMANTAN TENGAH	7230,10
22	KALIMANTAN SELATAN	1444,20
23	KALIMANTAN TIMUR	3786,50
24	KALIMANTAN UTARA	305,10
25	SULAWESI UTARA	3746,30
26	SULAWESI TENGAH	383,60
27	SULAWESI SELATAN	105,70
28	SULAWESI TENGGARA	106,10
29	GORONTALO	9,90
30	SULAWESI BARAT	386,20
31	MALUKU	23,60
32	MALUKU UTARA	386,20
33	PAPUA BARAT	88,10
34	PAPUA	345,10

Gambar 4. Kejanggalan yang Ditemukan Pada Dataset Tahun 2018

Dapat dilihat dari Gambar 4 jumlah provinsi yang masuk didalam baris berwarna kuning ini ada 5 provinsi, yaitu DKI Jakarta, DI Yogyakarta, Nusa Tenggara Barat, Sulawesi Utara, dan Maluku Utara. Kejanggalaan tersebut yaitu terdapat perulangan nilai jumlah produksi yang ditunjukkan oleh panah pada nilai produksi setiap provinsi dengan baris berwarna kuning. Oleh sebab itu, dilakukan pengkajian terhadap sumber dari *dataset* ini yaitu pada *website* tersebut tercantum sumber aslinya yaitu dari Direktorat Jenderal Perkebunan. Berikut Gambar 5 yang menunjukkan data asli produksi kelapa sawit selama 5 tahun terakhir sejak 2021 dari Direktorat Jenderal Perkebunan yang diunggah langsung di laman *website* Kementerian Pertanian RI.

No	Provinsi/Provinsi	Tahun					Perubahan tahun 2021 over 2018 (%)
		2017	2018	2019	2020	2021	
1	Aceh	0	0	0	0	0	0,00
2	Sumatera Utara	1.158.807	1.027.432	1.139.347	1.154.939	1.197.037	3,25
3	Sumatera Barat	1.432.972	1.346.294	1.378.164	1.313.191	1.375.120	6,41
4	Riau	3.114.822	3.448.220	3.873.228	3.884.314	3.272.749	-11,07
5	Jambi	20.004	20.000	20.744	20.000	20.000	0,00
6	Sulawesi Utara	1.640.959	1.807.275	2.144.402	2.222.989	2.138.223	7,19
7	Sulawesi Selatan	1.180.081	1.719.822	1.938.189	1.947.221	1.989.791	6,74
8	Kalimantan Tengah	622.048	622.019	618.007	644.247	629.402	-6,42
9	Maluku	803.022	1.047.728	1.022.000	1.063.424	1.221.493	15,82
10	Lampung	482.714	487.223	414.222	484.444	482.827	-14,09
11	DKI Jakarta	-	-	-	-	-	0,00
12	DI Yogyakarta	48.888	48.024	52.197	52.003	52.318	6,11
13	Nusa Tenggara Barat	52.881	58.428	81.022	57.422	58.121	-19,48
14	Nusa Tenggara Timur	-	-	-	-	-	0,00
15	Sumatera Selatan	-	-	-	-	-	0,00
16	Sumatera Tengah	-	-	-	-	-	0,00
17	Sumatera Barat	-	-	-	-	-	0,00
18	Sumatera Utara	-	-	-	-	-	0,00
19	Sumatera Selatan	-	-	-	-	-	0,00
20	Kalimantan Barat	1.794.120	3.090.888	3.222.244	3.471.407	3.222.993	89,82
21	Kalimantan Tengah	1.774.811	2.232.034	2.394.341	2.768.778	2.622.482	9,21
22	Kalimantan Selatan	1.822.721	1.424.228	1.582.187	1.581.147	1.924.378	74,74
23	Kalimantan Timur	2.842.710	3.788.477	3.888.889	3.824.221	3.924.244	9,48
24	Kalimantan Utara	214.222	222.122	221.122	221.122	221.122	0,00
25	Sulawesi Tengah	-	-	-	-	-	0,00
26	Sulawesi Tenggara	1.720	1.841	18.719	4.378	18.221	92,11
27	Sulawesi Selatan	487.428	487.427	487.427	487.427	487.427	-0,01
28	Sulawesi Utara	119.972	122.728	122.728	122.728	122.728	0,00
29	Sulawesi Barat	842.718	842.717	842.718	842.718	842.718	-0,02
30	Sulawesi Tenggara	84.427	101.118	101.118	101.118	101.118	19,16
31	Maluku	11.424	22.422	11.424	11.424	11.424	0,00
32	Maluku Utara	-	-	-	-	-	0,00
33	Irian Jaya	143.822	143.822	143.822	143.822	143.822	0,00
34	Irian Barat	143.822	143.822	143.822	143.822	143.822	0,00
Total		37.968.221	42.883.221	47.122.221	48.287.221	49.715.221	16,00

Gambar 5. Data Produksi Kelapa Sawit Publikasi *Website* Kementerian Pertanian (Sumber: pertanian.go.id)

Dapat dilihat pada Gambar 5 diatas pada data dari sumber *website* kementerian pertanian memperlihatkan bahwa kelima provinsi tidak memproduksi tanaman kelapa sawit pada tahun 2018. Berbeda dengan data yang didapat sebelumnya dari *website* bps.go.id yang menunjukkan bahwa kelima provinsi tersebut memiliki jumlah nilai produksi (dapat dilihat pada Tabel 1).

Selanjutnya, untuk memastikan validasi pada kondisi data ini dilakukan konfirmasi dengan pihak badan pusat statistik dengan cara menghubungi *website* Sistem Informasi Layanan Statistik (SILASTIK) yaitu *silastik.bps.go.id*, yang mana merupakan sistem informasi layanan untuk bertanya perihal data yang dipublikasi oleh BPS. Berikut ini Gambar 6 yang menunjukkan balasan ajuan konfirmasi validasi data yang dikirim ke *website* Sistem Informasi Layanan Statistik (SILASTIK).



Gambar 6. Balasan Ajuan Konfirmasi Validasi Data Dari SILASTIK

Dilihat dari Gambar 6 balasan yang dikirimkan oleh pihak Silastik menyatakan bahwa benar adanya kesalahan pada data yang tercantum di *website* BPS. Setelah dilakukan pengecekan ulang, kemudian data yang terdapat di *website* BPS tersebut diperbaiki sesuai dengan sumber yang dicantumkan yaitu dari Direktorat Jenderal Perkebunan.

Setelah proses konfirmasi terkait kejanggalan pada dataset didapat bahwa kelima provinsi tersebut tidak pernah memproduksi tanaman perkebunan kelapa sawit dari tahun 2008 s.d 2021. Berikut ini Gambar 7 yang menunjukkan daftar 9 provinsi yang tidak memproduksi tanaman perkebunan kelapa sawit.

No.	Provinsi	Produksi Tanaman Perkebunan Kelapa Sawit (Kilowon Ton)													
		2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
1	JAWA TENGAH	--	--	--	--	0,00	0,00	0,00	0,00	--	--	--	--	--	--
2	JAWA TIMUR	--	--	--	--	0,00	0,00	0,00	0,00	--	--	--	--	--	--
3	BALI	--	--	--	--	0,00	0,00	0,00	0,00	--	--	--	--	--	--
4	NUSA TENGGARA TIMUR	--	--	--	--	0,00	0,00	0,00	0,00	--	--	--	--	--	--
5	DKI JAKARTA	--	--	--	--	0,00	0,00	0,00	0,00	--	--	--	--	--	--
6	DI YOGYAKARTA	--	--	--	--	0,00	0,00	0,00	0,00	--	--	--	--	--	--
7	NUSA TENGGARA BARAT	--	--	--	--	0,00	0,00	0,00	0,00	--	--	--	--	--	--
8	SULAWESI UTARA	--	--	--	--	0,00	0,00	0,00	0,00	--	--	--	--	--	--
9	MALUKU UTARA	--	--	--	--	0,00	0,00	0,00	0,00	--	--	--	--	--	--

Gambar 7. Daftar Provinsi yang Tidak Memproduksi Kelapa Sawit

Sehingga, dengan ini ke sembilan provinsi yang tidak memproduksi tanaman perkebunan kelapa sawit tersebut akan dihapus agar nantinya tidak mempengaruhi nilai hasil *clustering*. Dataset penelitian ini memiliki 15 atribut yaitu atribut provinsi bertipe data *char*, lalu atribut produksi dalam rentang 14 tahun yaitu tahun 2008 s.d 2021 bertipe data number (*integer*). Berikut ini Gambar 8 yang menunjukkan struktur data pada *dataset* didalam penelitian.

```
> str(data1)
tibble [23 × 15] (S3: tbl_df/tbl/data.frame)
 $ PROVINSI: chr [1:23] "ACEH" "SUMATERA UTARA" "SUMATERA BARAT" "RIAU" ...
 $ 2008 : num [1:23] 565 3882 963 4813 1626 ...
 $ 2009 : num [1:23] 693 3862 896 5311 1300 ...
 $ 2010 : num [1:23] 616 3900 986 5496 1644 ...
 $ 2011 : num [1:23] 592 4011 927 5896 1773 ...
 $ 2012 : num [1:23] 655 3975 930 6385 1718 ...
 $ 2013 : num [1:23] 818 4549 1022 6647 1750 ...
 $ 2014 : num [1:23] 946 4870 925 6993 1774 ...
 $ 2015 : num [1:23] 896 5193 927 8060 1795 ...
 $ 2016 : num [1:23] 733 3984 1183 7668 1435 ...
 $ 2017 : num [1:23] 867 4852 1209 7591 1783 ...
 $ 2018 : num [1:23] 1037 5737 1248 8496 2691 ...
 $ 2019 : num [1:23] 1133 5647 1253 9513 2884 ...
 $ 2020 : num [1:23] 1135 5777 1312 9984 3023 ...
 $ 2021 : num [1:23] 1036 5311 1352 8629 2575 ...
```

Gambar 8. Struktur data pada *dataset*.

Selanjutnya, perlu analisis lebih lanjut untuk mengetahui nilai analisis statistika deskriptif dari setiap atribut yang ada dalam dataset. Berikut ini Gambar 9 yang menunjukkan hasil *outputnya*.

```

summary(data1)
# Province
# Length:22
# Class: character
# Mode: character

2008      2009      2010
min.    1  10.80 min.    1  0.00 min.    1  14.10
1st Qu.: 42.18 1st Qu.: 58.93 1st Qu.: 83.73
median : 414.40 median : 431.50 median : 553.33
mean    : 881.43 mean    : 872.29 mean    :1022.59
3rd Qu.:1103.47 3rd Qu.:1239.12 3rd Qu.:1332.47
max.    :14812.90 max.    :15111.40 max.    :15496.00
NA's    : 13      NA's    : 13      NA's    : 12

2011      2012      2013      2014
min.    1  15.20 min.    1  22.72 min.    1  14.74 min.    1  24.10
1st Qu.: 66.13 1st Qu.: 83.68 1st Qu.: 80.88 1st Qu.: 88.43
median : 548.40 median : 579.70 median : 460.09 median : 516.40
mean    :1089.81 mean    :1187.52 mean    :1137.58 mean    :1272.86
3rd Qu.:1344.22 3rd Qu.:1611.29 3rd Qu.:1449.74 3rd Qu.:1587.13
max.    :1893.50 max.    :6184.34 max.    :16647.00 max.    :16993.20
NA's    : 13      NA's    : 13      NA's    : 11      NA's    : 12

2015      2016      2017      2018
min.    1  0.10 min.    1  7.30 min.    1  1.5 min.    1  8.9
1st Qu.: 89.82 1st Qu.: 97.27 1st Qu.: 113.1 1st Qu.: 103.7
median : 478.70 median : 590.35 median : 527.3 median : 487.2
mean    :1294.38 mean    :1322.12 mean    :1387.6 mean    :1711.1
3rd Qu.:1618.47 3rd Qu.:1860.91 3rd Qu.:1793.0 3rd Qu.:1891.3
max.    :8039.80 max.    :7666.10 max.    :17991.2 max.    :18496.0
NA's    : 13      NA's    : 13      NA's    : 11      NA's    : 12

2019      2020      2021
min.    1  24.2 min.    1  5.0 min.    1  6.5
1st Qu.: 91.0 1st Qu.: 100.3 1st Qu.: 84.4
median : 437.7 median : 557.9 median : 724.4
mean    :1884.8 mean    :1931.9 mean    :1848.9
3rd Qu.:2884.4 3rd Qu.:3022.6 3rd Qu.:2575.1
max.    :9512.9 max.    :9984.3 max.    :9629.1
    
```

Gambar 9. Hasil Output Nilai Statistik Deskriptif Pada Dataset

Dari Gambar 9 dapat terlihat nilai *mean*, *median*, *quartil*, *min* dan *max* dari setiap masing-masing atribut. Selanjutnya, dapat dilihat bahwa ada keterangan nilai *missing values* (NA'S) di sembilan kolom atribut tahun produksi yaitu pada tahun 2008-2016.

2. Verify Data Quality

Setelah penjelasan pada bagian EDA dapat disimpulkan bahwa kualitas data sudah cukup baik. Namun, masih seperti pemecahan masalah untuk *missing value* (NA) agar nantinya tidak mengganggu proses hasil *cluster*. Lalu jika dilihat kembali, dalam dataset ini terdapat rentang nilai tiap data (*outliers*) yang cukup besar antara satu dengan lainnya. Sehingga, hal itu juga perlu dilakukan tambahan teknik normalisasi data pada tahap selanjutnya.

Data Preparation

1. Data Preprocessing

Penanganan untuk *missing value* yang ada di dalam dataset. Berikut ini Gambar 10 yang akan memperlihatkan *missing value* yang ada didalam *dataset*.

```

17 KALIMANTAN UTARA      NA      NA      NA      NA      NA      164.79      255.7      268.1      167.7
18 SULAWESI TENGAH      126.8      144.3      148.8      185.1      234.06      244.07      234.4      279.3      316.8
19 SULAWESI SELATAN      21.5      28.2      34.8      42.7      47.18      48.82      78.8      111.5      105.1
20 SULAWESI TENGGARA      10.8      2.5      15.2      16.2      28.86      71.28      71.0      72.4      85.4
21 GORONTALO            NA      NA      NA      NA      NA      NA      NA      NA      NA
22 SULAWESI BARAT      325.8      292.5      264.4      249.3      253.73      262.74      285.5      294.6      434.1
23 NALURU              NA      NA      NA      NA      NA      14.74      NA      NA      NA
24 PAPUA BARAT          40.0      57.7      66.2      34.1      93.72      93.72      74.0      74.0      188.9
25 PAPUA                49.5      68.7      136.4      123.1      85.46      85.46      94.0      95.5      226
    
```

Gambar 10. Missing Value Pada Dataset

Pada penelitian ini penanganan *missing value* menggunakan *input* nilai *median*. Berikut ini gambar 11 dan gambar 12 yang menunjukkan *source code* dan juga hasil *output* penanganan *missing value* dengan menggunakan nilai *median*.

```

25 # Mengganti NA menjadi nilai median
26 data1$2008 ~ifelse(is.na(data1$2008), median(data1$2008, na.rm=TRUE), data1$2008)
27 data1$2009 ~ifelse(is.na(data1$2009), median(data1$2009, na.rm=TRUE), data1$2009)
28 data1$2010 ~ifelse(is.na(data1$2010), median(data1$2010, na.rm=TRUE), data1$2010)
29 data1$2011 ~ifelse(is.na(data1$2011), median(data1$2011, na.rm=TRUE), data1$2011)
30 data1$2012 ~ifelse(is.na(data1$2012), median(data1$2012, na.rm=TRUE), data1$2012)
31 data1$2013 ~ifelse(is.na(data1$2013), median(data1$2013, na.rm=TRUE), data1$2013)
32 data1$2014 ~ifelse(is.na(data1$2014), median(data1$2014, na.rm=TRUE), data1$2014)
33 data1$2015 ~ifelse(is.na(data1$2015), median(data1$2015, na.rm=TRUE), data1$2015)
34 data1$2016 ~ifelse(is.na(data1$2016), median(data1$2016, na.rm=TRUE), data1$2016)
35 summary(is.na(data1))
36 view(data1)
    
```

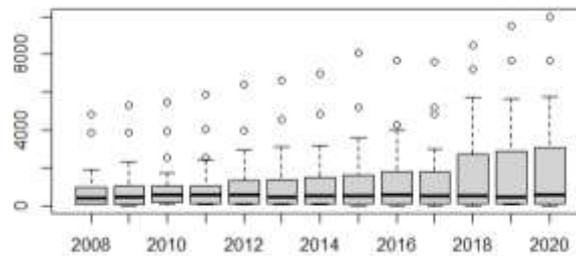
Gambar 11. Source Code Penanganan Missing Value Dengan Nilai Median

17	KALIMANTAN UTARA	414.6	491.5	553.35	548.4	579.70	164.73	259.7	268.1	167.70
18	SULAWESI TENGAH	126.6	144.3	145.80	185.1	234.08	244.07	254.4	275.3	316.80
19	SULAWESI SELATAN	21.5	28.2	34.90	42.7	47.18	49.82	78.9	111.5	105.10
20	SULAWESI TENGGARA	10.8	0.0	13.20	16.2	28.98	71.28	71.0	72.4	63.40
21	GORONTALO	414.6	491.5	553.35	548.4	579.70	466.09	516.6	51	500.35
22	SULAWESI BARAT	325.8	280.5	264.40	349.3	233.73	282.74	285.5	294.6	434.10
23	MALUKU	414.6	491.5	553.35	548.4	579.70	14.74	516.6	478.7	7.30
24	PAPUA BARAT	40.0	57.7	66.20	54.1	53.72	53.72	74.0	74.0	135.90
25	PAPUA	49.5	66.7	136.40	102.1	83.48	83.48	94.0	93.5	20.60

Gambar 12. Hasil Output Penanganan Missing Value Dengan Nilai Median

2. *Data Normalization*

Sebelum ketahap normalisasi menggunakan metode *min-max normalization*, langkah pertama adalah mengecek terlebih dahulu *outlier* pada data dengan menggunakan *boxplot*. Berikut Gambar 13 yang menunjukkan hasil visualisasi *outliers* pada data menggunakan diagram *boxplot*.



Gambar 13. Visualisasi *Outliers* Pada Data Menggunakan Diagram *Boxplot*

Dapat dilihat dari gambar 13 diatas, hasil visualisasi dari diagram *boxplot* tersebut yaitu adanya nilai *outliers* yang cukup tinggi di setiap tahunnya. Oleh karena itu, dilakukan tahap normalisasi menggunakan *min-max normalization* untuk menangani nilai *outlier* tersebut.

```

50 # min-max normalization
51 normalize = function(x) {
52   return ((x - min(x)) / (max(x) - min(x)))
53 }
54 data2 = as.data.frame(sapply(data1[2:14], normalize))
55 view(data2)
56

```

Gambar 14. Source Code Normalisasi Menggunakan *Min-Max Normalization*

	2008	2009	2010	2011	2012	2013
1	0.1153822127	0.130474075	0.1098889071	0.0981242454	0.102560247	0.122992327
2	0.8062386773	0.727190571	0.7087870994	0.6794721358	0.622665063	0.684398977
3	0.1980092872	0.168750235	0.1772743027	0.1551281397	0.145683166	0.153803219
4	1.0000000000	1.0000000000	1.0000000000	1.0000000000	1.0000000000	1.0000000000
5	0.3364846011	0.282392590	0.2973421624	0.2989473326	0.269132937	0.263219498
6	0.3916456698	0.435572542	0.4612816724	0.4085675901	0.390458827	0.404787122

Gambar 15. Hasil Output Normalisasi Menggunakan *Min-Max Normalization*

Dapat dilihat dari Gambar 15 yang menunjukkan hasil potongan gambar *output* diatas, bahwa nilai dari tiap *variabel* tahun berubah dalam bentuk rentang nilai 0-1. Dengan ini, masalah *outlier* sudah terselesaikan, sehingga data dapat lanjut ketahap selanjutnya yaitu tahap *modelling*.

Modelling

1. *Build Model with K-Medoid*

Algoritma *k-medoids* menggunakan metode partisi *clustering* untuk melakukan pengelompokan terhadap kumpulan n objek menjadi sejumlah k *cluster* (Pramessti et al. 2017). Oleh karena itu, diperlukan penentuan jumlah k *cluster* terbaik untuk

Berdasarkan dari Gambar 18 dapat dilihat representasi titik dan garis yang berhubungan antar *cluster* yaitu bahwa hasil plot tiap *dot cluster* tidak ada yang saling menimpa antar *clusternya*, maka dapat dikatakan *cluster* sudah ideal. Berikut ini Gambar 19 dan Gambar 20 yang menunjukkan *source code* dan hasil *output* berisi hasil tabel *cluster* beserta nama provinsinya.

```
78 # Melihat hasil clustering dalam bentuk tabel
79 data1$cluster = result$cluster
80 view(data1)
```

Gambar 19. Source Code Menampilkan Hasil Tabel Cluster

PROVINSI	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	cluster
1 ACEH	984.7	693.0	816.93	593.2	654.80	817.53	945.8	896.3	732.7	867.3	1037.4	1133.3	1134.8	1036.1	1
2 BANTEN	39.0	25.1	36.00	21.8	27.67	27.06	24.3	25.5	27.3	32.1	38.4	31.3	27.4	30.1	1
3 BENGKULU	983.2	736.0	706.00	827.1	800.02	767.05	766.8	747.5	750.2	849.7	1047.7	1032.1	1083.4	1152.7	1
4 GORONTALO	474.8	451.5	553.35	548.4	0.00	0.00	0.0	0.1	0.0	1.5	9.9	16.2	9.0	6.5	1
5 JAMBI	1628.8	1499.9	1844.10	1775.1	1718.29	1748.62	1775.7	1794.9	1435.1	1760.0	2091.5	2884.4	3022.8	2575.1	3
6 JAWA BARAT	12.8	20.7	16.40	22.1	22.72	32.64	33.0	33.0	32.8	42.4	46.0	33.2	33.1	32.8	1
9 SULAWESIA BARAT	1151.4	1054.7	1152.80	1443.8	1831.62	1764.07	1802.2	1483.1	1483.8	1018.0	1368.8	2192.3	2471.4	2008.8	3

Gambar 20. Hasil Output Tampilan Tabel Cluster Beserta Nama Provinsinya

Selanjutnya yaitu pemberian nama label. Pemberian nama label akan dilihat dari pendekatan nilai *mean* tiap *cluster* di atribut tahun produksinya. Berikut ini *source code* dan hasil *output* dari perhitungan nilai *mean cluster* pada setiap atribut tahun produksi.

```
82 # melihat deskripsi data setelah di clustering
83 mean_cluster = data1[2:14]%>%
84 mutate(cluster=result$cluster)%>%
85 group_by(cluster)%>%
86 summarise_all(mean)
87 print(mean_cluster)
88
```

Gambar 21. Source Code Perhitungan Nilai Mean Setiap Cluster

```
> print(mean_cluster)
# A tibble: 3 x 14
  cluster 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
  <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1 315. 343. 373. 371. 324. 326. 351. 326. 384. 414.
2 2 4348. 4587. 4698. 4953. 5180. 5598. 5932. 6626. 6826. 6222.
3 3 1255. 1480. 1608. 1808. 2060. 2142. 2219. 2389. 2635. 3010.
```

Gambar 22. Hasil Output Perhitungan Nilai Mean Setiap Cluster

Dapat dilihat dari gambar 22 bahwa nilai *mean* pada setiap *cluster 2* memiliki nilai produksi yang paling tinggi setiap tahunnya dibanding dengan *cluster* lain, sehingga provinsi yang termasuk *cluster* ini adalah provinsi dengan tingkat daerah produksi tinggi (*High*). Selanjutnya, nilai pada *cluster 3* memiliki nilai produksi lebih rendah dibanding *cluster 1*, sehingga provinsi yang termasuk *cluster* ini adalah produksi provinsi dengan daerah tingkat sedang (*Medium*), Kemudian, nilai pada *cluster 1* memiliki nilai paling rendah dibandingkan dengan *cluster* lainnya, sehingga provinsi yang termasuk *cluster* ini adalah produksi provinsi dengan daerah tingkat rendah (*Low*). Berikut ini Tabel 1 yang menjelaskan hasil *cluster* beserta label tingkatannya

Tabel 1. Hasil Cluster Setiap Provinsi Beserta Label Tingkat Produksinya

Provinsi	Cluster	Tingkat Produksi
Aceh	1	Rendah
Sumatera Utara	2	Tinggi
Sumatera Barat	1	Rendah
Riau	2	Tinggi
Jambi	3	Sedang
Sumatera Selatan	3	Sedang
Bengkulu	1	Rendah

Lampung	1	Rendah
Kep. Bangka Belitung	1	Rendah
Kep. Riau	1	Rendah
Jawa Barat	1	Rendah
Banten	1	Rendah
Kalimantan Barat	3	Sedang
Kalimantan Tengah	3	Sedang
Kalimantan Selatan	1	Rendah
Kalimantan Timur	3	Sedang
Kalimantan Utara	1	Rendah
Sulawesi Tengah	1	Rendah
Sulawesi Selatan	1	Rendah
Sulawesi Tenggara	1	Rendah
Gorontalo	1	Rendah
Sulawesi Barat	1	Rendah
Maluku	1	Rendah
Papua Barat	1	Rendah
Papua	1	Rendah

Dapat dilihat dari Tabel 1 diatas, bahwa *cluster* ke-1 sebanyak 18 provinsi dengan tingkat daerah produksi rendah (*Low*), Kemudian, *cluster* 2 sebanyak 2 provinsi dengan tingkat daerah produksi tinggi (*High*). Dan selanjutnya, *cluster* 3 sebanyak 5 provinsi dengan tingkat daerah produksi sedang (*Medium*).

Evaluation

Dalam penelitian ini proses evaluasi model menggunakan teknik evaluasi *silhouette coefficient*. Berikut ini Gambar 23 yang menunjukkan *source code* dan Gambar 24 yang menunjukkan hasil *output* evaluasi yang telah dilakukan.

```
65 # Evaluasi menggunakan teknik silhouette coefficient
66 fviz_silhouette(result)
67
```

Gambar 23. *Source Code* Evaluasi Menggunakan Teknik *Silhouette Coefficient*



Gambar 24. Hasil Evaluasi Menggunakan Teknik *Silhouette Coefficient*

Dilihat dari gambar grafik evaluasi tersebut dapat dilihat bahwa nilai rata-rata evaluasi *cluster* dengan *silhouette coefficient* sebesar sebesar 0,64. Nilai evaluasi tersebut menurut Rousseeuw (dalam Wira et al., 2019) masuk dalam kriteria *medium structure* (*structure cluster* yang baik).

Deployment

Tujuan visualisasi dalam bentuk pemetaan ini yaitu supaya hasil *cluster* dapat lebih mudah dilihat dan dipahami. Dimana, setiap wilayah provinsi di representasikan dengan warna yang berbeda sesuai dengan hasil *cluster* yang didapat pada tahap *modelling*. Adapun untuk daerah pada peta yang berwarna abu, merupakan daerah provinsi yang setiap tahunnya tidak memproduksi tanaman kelapa sawit. Berikut ini Gambar 25 yang menunjukkan hasil pemetaan menggunakan bantuan tools QGIS.



Gambar 25. Hasil Pemetaan Daerah Provinsi Menggunakan Software QGIS

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat diambil beberapa kesimpulan sebagai berikut ini.

1. Penerapan algoritma *k-medoids* pada penelitian ini *cluster* yang didapat sebanyak 3 *cluster*, yaitu pada *cluster* 2, ada sebanyak 2 provinsi yang masuk dalam *cluster* ini yaitu Riau dan Sumatera Utara yang termasuk dalam kategori provinsi dengan daerah produksi tinggi tanaman kelapa sawit (*High*). Selanjutnya, pada *cluster* 3 ada sebanyak 5 provinsi yang masuk dalam *cluster* ini yaitu Sumatera Selatan, Kalimantan Barat, Jambi, Kalimantan Tengah, dan Kalimantan Timur yang masuk dalam kategori provinsi dengan daerah produksi rendah tanaman kelapa sawit (*Medium*). Selanjutnya, pada *cluster* 1 ada sebanyak 18 provinsi yang masuk dalam *cluster* ini yaitu Gorontalo, Jawa Barat, Aceh, Sulawesi Barat, Banten, Sumatera Barat, Bengkulu, Lampung, Maluku, Kep. Bangka Belitung, Kep. Riau, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Papua Barat, dan Papua yang termasuk dalam kategori provinsi dengan daerah produksi rendah tanaman kelapa sawit (*Low*).
2. Hasil evaluasi pemetaan daerah produksi perkebunan kelapa sawit pada provinsi Indonesia menggunakan algoritma *k-medoids* dan metode evaluasi *silhouette coefficient*, dimana hasil rentang nilai *silhouette coefficient* pada perhitungan model menggunakan teknik ini yaitu sebesar 0,64. Sehingga, nilai rentang perhitungan model menggunakan *silhouette coefficient* pada penelitian ini termasuk dalam kriteria *medium structure* (*structure cluster* yang baik).

REFERENCES

- Buulolo, Efori. (2020). *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: Deepublish.
- Direktorat Jenderal Perkebunan Kementerian Pertanian RI. (2021). *Statistik Perkebunan Unggulan Nasional 2019-2020*. Jakarta: Penulis. Retrieved from <https://ditjenbun.pertanian.go.id/template/uploads/2021/04/BUKUSTATISTIK-PERKEBUNAN-2019-2021-OK.pdf>
- Effendi, H., Syahrial, A., Prayoga, S., & Hidayat, W. D. (2021). Penerapan Metode K-Means Clustering Untuk Pengelompokan Lahan Sawit Produktif Pada PT Kasih Agro Mandiri. *Teknomatika*, 11(02), 117-126.
- Ginantara, Ni Luh Wiwik Sri Rahayu, Fatimah Nur Arifah, Anggi Hadi Wijaya, Ri Sabti Septarini, Faried Effendy, and Edi Surya Negara. (2021). *Data Mining Dan Penerapan Algoritma*. edited by R. W. & J. Simarmata. Yayasan Kita Menulis.
- Haryadi, D. (2021). Penerapan Algoritma K-Means Clustering Pada Produksi Perkebunan Kelapa Sawit Menurut Provinsi. *Journal of ICT (Informatics and Communication Technology)*, 3(1), 50-64.
- Mauritsius, Tuga, and Faisal Binsari. 2020. "Cross-Industry Standard Process for Data Mining (CRISP-DM)." *Binus.Ac.Id*. Retrieved March 4, 2022 (<https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for-data-mining-crisp-dm/>).
- Muningsih, E., & Kiswati, S. (2018). Sistem aplikasi berbasis optimasi metode elbow untuk penentuan clustering pelanggan. *Joutica*, 3(1), 117-124.
- Pramesti, Dyang Falila, Lahan, M. Tanzil Furqon, and Candra Dewi. (2017). "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 1(9):723–32.
- Wira, Bagus., Alexius Endy. Budiando, and Anggri S. Wiguna. (2019). "Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi." *Jurnal Terapan Sains & Teknologi* 1(3):54–69.
- www.kompas.com.(2022, 31 Januari). RI Penghasil Sawit Terbesar Dunia, tapi Harganya Diatur dari Malaysia. Diakses pada 03 Februari 2022, dari <https://money.kompas.com/read/2022/01/30/062749426/ri-penghasil-sawit-terbesar-dunia-tapi-harganya-diatur-dari-malaysia?page=all>
- www.sawitindonesia.com. (2021, 25 Desember). Tiga Menteri Jokowi Akui Ketangguhan Sawit. Diakses pada 24 Februari 2022, dari <https://sawitindonesia.com/tiga-menteri-jokowi-akui-ketangguhan-sawit/>