



Tersedia Online : <http://e-journals.unmul.ac.id/>

ADOPSI TEKNOLOGI DAN SISTEM INFORMASI (ATASI)

Alamat Jurnal : <http://e-journals2.unmul.ac.id/index.php/atasi/index>



Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara

Ramadhana ^{1)*}, Islamiyah ²⁾, Amin Padmo Azam Masa ³⁾

Program Studi Sistem Informasi, Fakultas Teknik, Universitas Mulawarman

E-Mail : rdana8707@gmail.com ¹⁾; islamiyahunmul@gmail.com ²⁾; aminpadmo@ft.unmul.ac.3 ³⁾;

ARTICLE INFO

Article history:

Received : 16 May 2023

Revised : 06 June 2023

Accepted : 10 June 2023

Available online : 26 June 2023

Keywords:

Clustering,
K-Means,
Data Mining,
Export,
Coal

Kata Kunci :

Clustering,
K-Means,
Data Mining,
Ekspor,
Batubara

APA style in citing this article:

Ramadhana, Islamiyah, & Masa, A. P. A. (2023). *Ramadhana Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara*. *Adopsi Teknologi Dan Sistem Informasi (ATASI)*, 2(1), 35 - 42.
<https://doi.org/10.30872/atasi.v2i1.595>

ABSTRACT

Indonesia is a country that includes exporters to developed countries and developing countries. This study discusses the Application of Data Mining in Coal Exports of Destination Countries Using the K-Means Clustering Method. The data source for this research was collected based on export-import information documents issued by the Directorate General of Customs and Excise. The data used in this study is data for 2012-2021 which consists of 10 variables used (1) total export net weight (tons). The data will be processed by clustering into 3 clusters, namely the high export level cluster, the medium export level cluster, and the low export level cluster. The high export rate cluster data center is 94681.9, the medium export level cluster data center is 17272.4, and the low export rate cluster data center is 77.2. In order to obtain an assessment using the Euclidean Distance and Manhattan Distance calculations with the same clustering results based on the coal export index with 4 cluster countries with high export levels namely India, China, Japan and South Korea and 5 countries with moderate export levels including Taiwan, Malaysia, the Philippines, Thailand, and other countries. In cluster countries the level of exports is low, namely Hong Kong and Spain. This can be input for the government, a country that is the highest priority in coal export activities based on the clusters that have been carried out.

ABSTRAK

Indonesia adalah negara yang termasuk pengeksport ke negara-negara maju dan berkembang. Penelitian ini membahas tentang Penerapan *Data mining* Pada Ekspor Batubara Negara Tujuan Menggunakan Metode *K-Means Clustering*. Data penelitian ini dikumpulkan berdasarkan dokumen-dokumen keterangan ekspor impor yang dihasilkan oleh Direktorat Jenderal Bea dan Cukai. Data yang digunakan dalam penelitian ini adalah data dari tahun 2012- 2021 yang terdiri dari 10 Variabel yang digunakan (1) jumlah ekspor berat bersih (ton). Data akan diolah dengan melakukan *clustering* dalam 3 *cluster* yaitu *cluster* tingkat ekspor tinggi, *cluster* tingkat ekspor sedang, dan *cluster* tingkat ekspor rendah. *Centroid* data untuk *cluster* tingkat ekspor tinggi 94681,9, *Centroid* data untuk *cluster* tingkat ekspor sedang 17272,4, dan *Centroid* data untuk *centroid* data untuk *cluster* tingkat ekspor rendah 77,2. Sehingga diperoleh penilaian menggunakan perhitungan jarak *Euclidean Distance* dan *Manhattan Distance* dengan hasil *clustering* yang sama berdasarkan indeks ekspor batubara dengan 4 negara *cluster* tingkat ekspor tinggi diantaranya adalah negara India, Tiongkok, Jepang, dan Korea Selatan serta 5 negara dengan tingkat ekspor sedang diantaranya Taiwan, Malaysia, Philipina, Thailand, dan negara lainnya. Pada negara *cluster* tingkat ekspor rendah yakni Hongkong dan Spanyol. Hal ini dapat menjadi masukan kepada pemerintah, negara yang menjadi prioritas tertinggi pada kegiatan ekspor batubara berdasarkan *cluster* yang telah dilakukan.

2023 ATASI: Adopsi Teknologi dan Sistem Informasi with CC BY NC SA license.

*) Corresponding Author

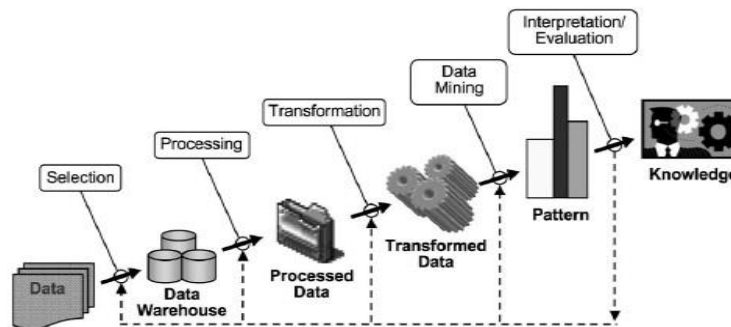
<https://doi.org/10.30872/atasi.v2i1.595>

1. PENDAHULUAN

Perdagangan internasional atau bisnis internasional antar perusahaan dilaksanakan dengan kontrak penjualan. Ke Berdasarkan data ekspor batubara di Indonesia pada tahun 2012 sampai dengan tahun 2021. Berdasarkan data yang bersumber dari badan pusat statistik pada halaman web <https://www.bps.go.id/> negara India menjadi urutan negara pertama dengan total pendapatan harga ekspor tertinggi dengan total 1.064.556,3 ton dan negara terendah terletak pada Spanyol dengan total 30.082,7 ton. Berdasarkan data tersebut *centroid* data untuk *cluster* tingkat ekspor tinggi sebesar 1.064.556,3, *centroid* data untuk *cluster* tingkat ekspor rendah sebesar 30.082,7 (Fadilah, 2018).

Indonesia merupakan salah satu negara yang memiliki sumber daya alam batubara yang melimpah. Batubara banyak dibutuhkan berbagai negara seluruh dunia termasuk Indonesia untuk digunakan sebagai bahan utama pembangkit listrik. Berdasarkan data dalam sepuluh tahun terakhir, pembangkit listrik tetap menjadi konsumen batubara terbesar di Indonesia. Perusahaan batubara di Indonesia sejak tahun 2005 merupakan salah satu produsen dan eksportir batubara terbesar di dunia, melampaui produksi Australia yang ditunjukkan oleh ekspor batubara dari Indonesia (Investment, 2018).

Data mining digunakan untuk membantu perusahaan mendapatkan informasi yang mereka butuhkan berdasarkan data yang ada. Penerapan *data mining* dalam hal ekspor batubara dapat dilakukan dengan membagi data tahun 2012 sampai dengan tahun 2021 menjadi beberapa kategori atau *cluster*. Hasil setiap kelompok yang berisi data yang sama dapat dianalisis pada kelompok tinggi, sedang dan rendah. KDD (*knowledge discovery in database*) adalah proses komputasi yang menggunakan algoritma matematika yang dirancang untuk mengekstrak data dan menghitung probabilitas kemungkinan tindakan dimasa depan (Sofiyana & Azkiya, 2022).



Gambar 1. Tahapan KDD (Sofiyana & Azkiya, 2022)

2. TINJAUAN PUSAKA

Clustering adalah suatu teknik pengelompokan data dengan cara memisahkan data menjadi beberapa kelompok menurut beberapa karakteristik yang diinginkan yang identitas kelompoknya masing-masing data tidak diketahui. Dengan pengelompokan ini diharapkan dapat mengetahui kelompok data dan kemudian diberikan identitas sesuai dengan permasalahan yang dihadapi (Gustientiedina et al., 2019)

K-means clustering pertama kali diperkenalkan oleh Hartigan pada tahun 1975. Metode ini sangat penting terutama untuk volume data yang besar karena akurasi yang lebih baik. *K-means clustering* adalah teknik pengelompokan data non-hierarkis yang bertujuan untuk mengubah atau membagi data menjadi beberapa kelompok. Untuk melakukan algoritma *clustering k-means*, data dibagi menjadi *k cluster* yang telah ditentukan (Mauser & Hartigan, 1977).

menjelaskan tahap-tahap menentukan *clustering* dengan metode *K-means Clustering* yang memiliki beberapa tahapan sebagai berikut (Pakpahan et al., 2022):

1. Menentukan data produksi komoditas data dan atribut-atribut yang akan digunakan
2. Melakukan normalisasi data ke rentang 0 sampai 1.
3. Penetapan jumlah *cluster*, pada penelitian ini berjumlah 3 *cluster* dari rendah, sedang, dan tinggi.
4. Menentukan *centroid* masing-masing *cluster*.
5. Menghitung jarak antara setiap titik data dan *Centroid* menggunakan perhitungan jarak *Euclidean Distance* dan *Manhattan distance*.

Perhitungan jarak *Euclidean Distance* menggunakan persamaan (1).

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_i - y_i)^2} \dots \dots \dots (1)$$

Keterangan :
d = Jarak Antara X Dan Y
x = Data Pusat *cluster*
y = Data Pada Atribut

- i = Setiap data
- n = Jumlah data
- x_i = Data pada pusat *cluster* ke- i
- y_i = Data pada setiap data ke- i

Perhitungan jarak *Manhattan Distance* menggunakan persamaan (2).

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \dots \dots \dots (2)$$

Keterangan :

- d = Jarak antara x dan y
- x = Data pusat *cluster*
- y = Data pada atribut

Perhitungan *cluster* dapat dilakukan dengan persamaan (3).

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x} \dots \dots \dots (3)$$

Keterangan :

- X_1 = Nilai data *record* ke-1, dst
- $\sum x$ = Jumlah data *record*

Pengelompokan object untuk menentukan anggota *cluster* adalah dengan memperhitungkan jarak minimum objek.

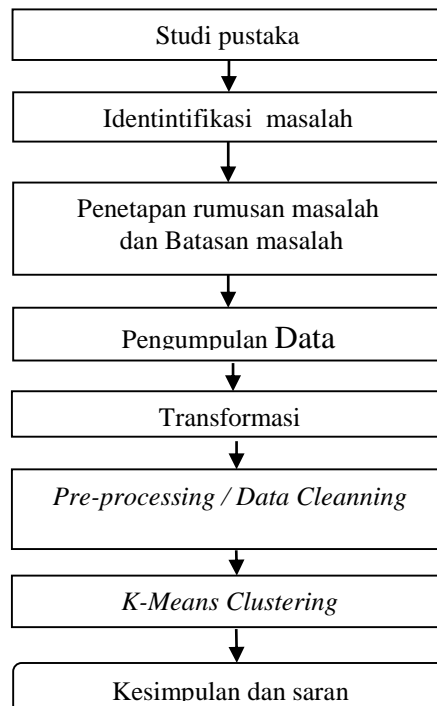
$$Z = \frac{x - \tilde{x}}{\sigma} \dots \dots \dots (4).$$

Keterangan

- Z = Nilai standar
- X = Data mentah
- \tilde{x} = Nilai rata rata
- σ = Simpang baku

Tujuan dari penelitian ini adalah mengetahui negara yang akan menjadi negara prioritas ekspor batubara dan mengetahui tingkat laba tertinggi dari setiap negara. Pengelompokan tersebut dapat menggunakan metode pengelompokan dengan algoritma *K-means Clustering*.

3. METODE PENELITIAN



Gambar 2. Tahapan Penelitian

Proses *data mining* pada data ekspor batubara dilakukan beberapa tahapan diantaranya:

1. Pengumpulan data

Teknik pengumpulan data yang digunakan diantaranya observasi. observasi atau pengamatan dilakukan dengan mengumpulkan data mengenai ekspor batubara. Data ekspor batubara secara umum terdapat pada Badan Pusat Statistik. Berdasarkan pengamatan yang telah dilakukan pada Badan Pusat Statistik, data ekspor baru bara hanya diolah dengan statistik penjumlahan ekspor batubara. Data yang akan digunakan untuk mengelompokkan adalah data produksi yang diperoleh dari data ekspor batubara menurut negara tujuan utama pada tahun 2012-2021 yang terkumpul sebanyak 110 data harga jual.

Tabel 1. Jumlah Ekspor Batubara Tahun 2012-2021 (*Indo_8_1659341892 (3), 2022*)

Negara Tujuan	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
India	94681,9	116824	134452	123842	95110,5	98553,5	110378	121693	98243,3	70779,2
Tiongkok	81528,9	89777,8	49782	36684,5	50961,1	48167,4	48135,7	65670,5	62492,5	108487
Jepang	35068,2	37711,5	35579,3	32503,5	33037,8	31421,4	28722,9	28436,4	26965,1	22978,4
Korea Selatan	37899,1	36149,6	35574,1	33037,3	34943,2	38075,1	37150,9	29550	24831,9	21011,2
Taiwan	28692,4	27947,2	27018,3	24088,1	20289,5	18187,7	17935,1	19061,2	17603	16291,6
Malaysia	16034,5	17120,6	14452,5	16567,5	17272,4	21189,9	22045,4	25323,5	26706,8	25497,2
Philipina	11621,1	14508,8	15021,3	15811,3	17503,4	18977,9	22595	27450,8	28060,9	30085,8
Thailand	14676	14258	16196,1	17729,5	16439	16374,7	19964,1	17600,4	16624,8	15150,7
Hongkong	11789,5	12875,6	12513,5	9414,7	9423,9	8449,8	9028,4	7876,8	3863,5	5269,9
Spanyol	5704,8	4078	4071,5	4826,5	4944	3232,2	2463,9	684,6	0	77,2
Lainnya	9808,1	10133,6	11642,4	13882,6	11405	16468,8	24704,7	31589,1	36155,8	29824,8

2. Transformasi data

Data ekspor batubara dapat diolah dengan menggunakan metode *K-Means Clustering*, dengan syarat data yang berjenis inisialisasi seperti nama negara harus diinisialisasikan terlebih dahulu dalam bentuk singkatan.

3. Pre-processing

Implementasi *pre-processing* adalah dengan melakukan *data cleaning* pada data pada tahun 2020 pada negara Spanyol dengan data sebesar 0 ton. Kasus tersebut akan digunakan 1 buah teknik *data cleaning* untuk membandingkan algoritma teknik *data cleaning* dengan nilai *rulesnya* mendekati benar. Hasil dari proses *Data Cleaning* memperoleh data 4.944,0 yang dimana data ini akan mengisi kekosongan data pada tahun 2020 untuk negara Spanyol. Untuk data ekspor batu bara 2012 – 2021 dapat dilihat pada tabel 1.

4. K-Means Clustering

Proses ini memiliki beberapa tahapan diantaranya agar mendukung proses perhitungan *K-Means Clustering*:

1. Tahap pertama mencari nilai rata-rata dari jumlah data ekspor batubara yang dimiliki.
2. Tahap kedua mencari nilai simpang baku menggunakan hasil perhitungan rata-rata dan data ekspor batubara.
3. Tahap ketiga setelah memperoleh nilai simpang baku kemudian melakukan perhitungan normalisasi data.
4. Tahap keempat melakukan perhitungan jarak menggunakan metode perhitungan jarak menggunakan *Euclidean Distance* dan *Manhattan Distance*.

4. HASIL DAN PEMBAHASAN

Data yang diolah dalam penelitian ini merupakan sampel yang diambil dari <https://www.bps.go.id>. Data sampel yang akan diuji cobakan terdiri dari 10 negara. Selanjutnya dicoba mengelompokkan data diatas menjadi 3 kelompok. Dengan menggunakan algoritma Kmeans, berikut langkah-langkah penyelesaiannya:

1. Menentukan Jumlah *cluster*, jumlah *cluster* merupakan jumlah kelompok yan akan dihasilkan. Dalam penelitian ini jumlah *cluster* yang akan digunakan adalah sebanyak 3 *cluster*.
2. Menentukan Centroid awal, Centroid awal diperoleh secara acak. Centroid awal merupakan titik pusat *cluster* pertama. Centroid data awal dari penelitian ini adalah:

$$C1 = 2,163$$

$$C2 = -0,525$$

$$C3 = -0,878$$

Tabel 2. *Preprocessing dan Data Cleaning*

Negara Tujuan	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
IND	94681,9	116824	134452	123842	95110,5	98553,5	110378	121693	98243,3	70779,2
RTT	81528,9	89777,8	49782	36684,5	50961,1	48167,4	48135,7	65670,5	62492,5	108487
JPN	35068,2	37711,5	35579,3	32503,5	33037,8	31421,4	28722,9	28436,4	26965,1	22978,4
KOR	37899,1	36149,6	35574,1	33037,3	34943,2	38075,1	37150,9	29550	24831,9	21011,2
TPE	28692,4	27947,2	27018,3	24088,1	20289,5	18187,7	17935,1	19061,2	17603	16291,6
MAS	16034,5	17120,6	14452,5	16567,5	17272,4	21189,9	22045,4	25323,5	26706,8	25497,2
PHI	11621,1	14508,8	15021,3	15811,3	17503,4	18977,9	22595	27450,8	28060,9	30085,8
THA	14676	14258	16196,1	17729,5	16439	16374,7	19964,1	17600,4	16624,8	15150,7
HKG	11789,5	12875,6	12513,5	9414,7	9423,9	8449,8	9028,4	7876,8	3863,5	5269,9
ESP	5704,8	4078	4071,5	4826,5	4944	3232,2	2463,9	684,6	4.944,0	77,2
LVN	9808,1	10133,6	11642,4	13882,6	11405	16468,8	24704,7	31589,1	36155,8	29824,8

3. Menghitung jarak setiap data yang ada terhadap setiap pusat *cluster*. Berikut perhitungannya:
Euclidean Distance yang dilakukan terdiri dari 4 proses iterasi. Pada iterasi pertama dilakukan perhitungan jarak data jumlah ekspor dari negara ke-1 ke pusat cluster pertama (C1) dengan menggunakan persamaan 2.1 sebagai iterasi pertama dilakukan berikut:

$$d_{11} = \sqrt{(2,1634 - 2,1634)^2 + (2,9204 - 2,9204)^2 + \dots + (1,3463 - 1,3463)^2}$$

$$d_{11} = 0$$

$$d_{12} = \sqrt{(2,1634 - (-0,5253))^2 + (2,9204 - (-0,4882))^2 + \dots + (1,3463 - (-0,2018))^2}$$

$$d_{12} = 9,568551988$$

$$d_{13} = \sqrt{(2,1634 - (-0,8785))^2 + (2,9204 - (-0,9341))^2 + \dots + (1,3463 - (-1,0708))^2}$$

$$d_{13} = 11,287$$

Adapun hasil dari perhitungan dari keseluruhan data terhadap tiap pusat *cluster* awal disajikan pada Tabel 3.

Tabel 3. Jarak Data Ke *Centroid* Iterasi Pertama *Euclidean Distance*

No	Negara Tujuan	Cluster1	Cluster2	Cluster3	Jarak Minimal	Kluster
1	IND	0	9,568552	11,28712	0	C1
2	RTT	5,908661	5,247603	6,982012	5,247603	C2
3	JPN	8,305524	1,490425	3,020443	1,490425	C2
4	KOR	8,14431	1,683421	3,213949	1,683421	C2
5	TPE	9,322177	0,926052	2,009633	0,926052	C2
6	MAS	9,568552	0	1,905716	0	C2
7	PHI	9,598418	0,264517	1,977802	0,264517	C2
8	THA	9,901036	0,602853	1,433662	0,602853	C2
9	HKG	10,67634	1,417654	0,663872	0,663872	C2
10	ESP	11,28712	1,905716	0	0	C3
11	LVN	9,699545	0,605572	2,084792	0,605572	C2

Perhitungan jarak menggunakan metode *Manhattan distance* pada iterasi pertama menggunakan titik *centroid* pada Tabel 4.1 sebelumnya. Pada iterasi pertama dilakukan perhitungan jarak data jumlah data ekspor batubara ke pusat *cluster* pertama (C1).

$$d_{11} = |2,1634 - 2,1634| + |2,9204 - 2,9204| + \dots + |1,3463 - 1,3463|$$

$$d_{11} = 0$$

$$d_{12} = |2,1634 - (-0,5253)| + |2,9204 - (-0,4882)| + \dots + |1,3463 - (-0,2018)|$$

$$d_{12} = 29,481$$

$$d_{13} = |2,1634 - (-0,8785)| + |2,9204 - (-0,9341)| + \dots + |1,3463 - (-1,0708)|$$

$$d_{13} = 35,197$$

Adapun hasil dari perhitungan dari keseluruhan data terhadap tiap pusat Cluster awal disajikan pada Tabel 4.

Tabel 4. Jarak Data Ke *Centroid* Iterasi Pertama *Euclidean Distance*

No	Negara tujuan	Cluster1	Cluster2	Cluster3	Jarak Minimal	Kluster
1	IND	0	29,4815	35,1972	0	C1
2	RTT	14,4569	15,0246	20,7403	14,4569	C1
3	JPN	25,7134	3,7681	9,4838	3,7681	C2
4	KOR	25,1734	4,3081	10,0238	4,3081	C2
5	TPE	28,9718	0,5097	6,2254	0,5097	C2
6	MAS	29,4815	0	5,7157	0	C2
7	PHI	29,501	0,2823	5,6962	0,2823	C2
8	THA	30,7532	-1,1787	4,444	-1,1787	C2
9	HKG	33,3002	-3,5285	1,897	-3,5285	C2
10	ESP	35,1972	-5,0093	0	-5,0093	C2
11	LNY	29,707	0,2003	5,4902	0,2003	C2

- Setelah semua data ditempatkan kedalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota ada pada *cluster* tersebut. Setelah didapat titik pusat baru dari tiap *cluster*, hitung kembali data dengan pusat *cluster* yang baru sampai didapat pola terakhir yang sudah tidak berpindah, dalam penelitian ini. Data dihitung ulang sampai iterasi ke 4, dimana setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* lainnya.
- Menghitung jarak data ke Centroid yang baru untuk iterasi berikutnya. Dengan menggunakan nilai Centroid yang baru, jarak ke masing-masing Centroid dihitung sampai group baru sama dengan group lama atau grup sebelumnya.
- Menghitung jarak data ke Centroid yang baru untuk iterasi berikutnya. Dengan menggunakan nilai Centroid yang baru, jarak ke masing-masing Centroid dihitung sampai grup baru sama dengan group lama atau group sebelumnya.

Tabel 5. Hasil iterasi ke empat pada metode *Euclidean Distance*

No	Negara Tujuan	Cluster1	Cluster2	Cluster3	Jarak Minimal	Kluster Literasi 3	Kluster Baru
1	IND	8,3231	9,9416	10,6763	8,32315	C1	C1
2	RTT	4,3503	5,7143	6,41167	4,35034	C1	C1
3	JPN	0,0534	1,7200	2,40941	0,05340	C1	C1
4	KOR	0,3999	1,90068	2,60139	0,39999	C1	C1
5	TPE	1,0404	0,85323	1,39545	0,85323	C2	C2
6	MAS	1,4658	0,59946	1,41765	0,59946	C2	C2
7	PHI	1,6055	0,74516	1,52681	0,74516	C2	C2
8	THA	1,6584	0,07680	0,89115	0,07680	C2	C2
9	HKG	2,3910	0,86760	0,00000	0,00000	C3	C3
10	ESP	3,0004	1,40366	0,66387	0,66387	C3	C3
11	LNY	1,8983	1,02342	1,72775	1,02342	C2	C2

Tabel 6. Hasil iterasi ke empat pada metode Manhattan Distance

No	Negara tujuan	Cluster1	Cluster2	Cluster3	Jarak	Kluster	Kluster iterasi ke 3
1	IND	25,767	30,858	33,3002	25,767	C1	C1
2	RTT	11,3101	16,401	18,8433	11,310	C1	C1
3	JPN	0,0536	5,1454	7,5868	0,0536	C1	C1
4	KOR	0,5936	5,6854	8,1268	0,5936	C1	C1
5	TPE	3,2048	1,887	4,3284	1,887	C2	C2
6	MAS	3,7145	1,3773	3,8187	1,3773	C2	C2
7	PHI	3,734	1,3578	3,7992	1,3578	C2	C2
8	THA	4,9862	0,1056	2,547	0,1056	C2	C2
9	HKG	7,5332	2,4414	0	0	C3	C3
10	ESP	9,4302	4,3384	1,897	1,897	C3	C3
11	LVN	3,94	1,1518	3,5932	1,1518	C2	C2

Kemudian dilakukan denormalisasi agar nilai data yang sebelumnya dilakukan normalisasi menjadi kembali seperti sebelumnya. Karena data perwakilan yang digunakan adalah data ke-1 yaitu negara india maka:

$$\begin{aligned} denor_{11} &= (2,163435653 * 29250,6) + 31400,10909 \\ denor_{12} &= (2,920397903 * 29250,6) + 31400,10909 \\ denor_{1\dots} &= \dots \\ denor_{110} &= (1,346266 * 29250,6) + 31400,10909 \end{aligned}$$

Metode pengukuran jarak yang digunakan dalam penelitian ini adalah *Sum of Squared Error* (SSE) yang akan memberikan informasi *error* jarak data ke *Centroid*. Dalam penelitian ini, hasil *clustering* metode pengukuran yaitu *Euclidean Distance* dan *Manhattan Distance* yang dibuat dengan tiga *cluster*. Nilai SSE dapat dilihat di Tabel 7.

Tabel 7. SSE 2 cluster

Cluster	Euclidean distance	Manhattan distance
K=2	140,2637989	1373,313368
K=3	238,1321014	2330,18749

Berdasarkan Tabel 7 dapat dilihat bahwa nilai SSE *Euclidean Distance* dengan tiga *cluster* adalah 238,1321014 dan dengan nilai dua *cluster* adalah 140,2637989. Pada nilai SSE *Manhattan Distance* dengan 3 *cluster* memperoleh nilai sebesar 2330,18749 dan dengan nilai dua *cluster* adalah 1373,313368.

5. KESIMPULAN

Berdasarkan dari hasil penelitian ekspor batubara menggunakan Metode Algoritma *K-Means Clustering* dapat disimpulkan hasil pengujian akurasi menggunakan metode SSE pada perhitungan jarak dengan 3 *cluster* pada *Euclidean Distance* diperoleh 238,1321014 dan pada 2 *cluster* diperoleh 140,2637989. Pada metode *Manhattan Distance* dengan 3 *cluster* diperoleh 2330,18749 dan pada 2 *cluster* diperoleh 1373,313368. Berdasarkan perbandingan jarak yang paling ideal dengan objek penelitian ekspor batubara, karena memiliki selisih nilai yang paling kecil adalah *Euclidean Distance* dimana metode ini selisihnya lebih kecil dibandingkan *Manhattan Distance*. Hasil pengelompokan diperoleh 3 *cluster* negara tujuan ekspor batubara yaitu: *Cluster I* = India, Tiongkok, Jepang, dan Korea Selatan. *Cluster II* = Taiwan, Malaysia, Philipina, Thailand, dan negara lainnya. *Cluster III* = Hongkong dan Spanyol. Berdasarkan pengelompokan negara, maka dapat dilakukan penentuan negara prioritas untuk mendapatkan laba tertinggi yaitu secara beturut-turut: negara India, Tiongkok, Korea Selatan, Jepang, Taiwan, Malaysia, Philipina, Negara Lainnya, Thailand, Hongkong, dan negara Spanyol. Pengelompokan negara tersebut juga dihasilkan negara prioritas diantaranya negara India, Tiongkok, dan Korea Selatan.

6. DAFTAR PUSTAKA

Afifi, W., Nastiti, D. R., & Aini, Q. (2020). Clustering K-Means Pada Data Ekspor (Studi Kasus: Pt. Gaikindo). *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 11(1), 45–50. <https://doi.org/10.24176/simet.v11i1.3568>

- Afin, A. P., & Kiono, B. F. T. (2021). Potensi Energi Batubara serta Pemanfaatan dan Teknologinya di Indonesia Tahun 2020 – 2050 : Gasifikasi Batubara. *Jurnal Energi Baru Dan Terbarukan*, 2(2), 144–122. <https://doi.org/10.14710/jebt.2021.11429>
- Almayda, A., & Saepudin, S. (2021). Penerapan *Data mining K-Means Clustering* Untuk Mengelompokkan Berbagai Jenis Merek Smartphone. *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 241–249.
- Amin, F., Anggraeni, D. S., & Aini, Q. (2022). Penerapan Metode *K-Means* dalam Penjualan Produk Souq.Com. *Applied Information System and Management (AISM)*, 5(1), 7–14. <https://doi.org/10.15408/aism.v5i1.22534>
- Anjelita, M., Perdana Windarto, A., & Wanto, A. (2019). Seminar Nasional Sains & Teknologi Informasi (SENSASI) Analisis Metode K-Means pada Kasus Ekspor Barang Perhiasan dan Barang Berharga Berdasarkan Negara Tujuan. *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, 476–482. <http://prosiding.seminar-id.com/index.php/sensasi/issue/archivePage%7C476>
- Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis *K-Means Clustering* pada Data Sepeda Motor. *INFORMAL: Informatics Journal*, 5(1), 10. <https://doi.org/10.19184/isj.v5i1.17071>
- Fadilah, F. (2018, april 04). Mengenal Ekspor Impor: Pengertian, Tujuan, Manfaat dan Komoditasnya. (Gamedia) Retrieved october 20, 2022, from <https://www.gamedia.com/literasi/ekspor-impor/>
- Galih Whendasmoro, R. (2022). Analisis Penerapan Normalisasi Data Dengan Menggunakan *Z-Score* Pada Kinerja Algoritma K-NN. *Jurnal Riset Komputer*, 9(4), 2407–389. <https://doi.org/10.30865/jurikom.v9i4.4526>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma *K-Means* Untuk *Clustering* Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Hajar, S., Novany, A. A., Windarto, A. P., Wanto, A., & Irawan, E. (2020). Penerapan *K-Means Clustering* pada ekspor minyak kelapa sawit menurut negara tujuan. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) 2020*, 314–318.
- Haryadi, H. (2021). Pengelolaan sumberdaya batubara Indonesia dan prospeknya dalam pasar global dengan analisis SWOT. *Jurnal Teknologi Mineral Dan Batubara*, 17(2), 107–122. <https://doi.org/10.30556/jtmb.vol17.no2.2021.1073>
- Haviluddin, H., Patandianan, S. J., Putra, G. M., Puspitasari, N., & Pakpahan, H. S. (2021). Implementasi Metode *K-Means* Untuk Pengelompokkan Rekomendasi Tugas Akhir. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 16(1), 13. <https://doi.org/10.30872/jim.v16i1.5182>
- Indo_8_1659341892 (3), (2022) (testimony of Badan Pusat Statistik). <https://www.bps.go.id/statictable/2014/09/08/1034/ekspor-batu-bara-menurut-negara-tujuan-utama-2012-2021.html>
- Investment, I. (2018, April 05). Komoditas Batubara. Retrieved from Invesment Indonesia: <https://www.indonesia-investments.com/id/bisnis/komoditas/batu-bara/item236?>
- Mausser, G. A., & Hartigan, J. A. (1977). *Clustering Algorithms*. In *Journal of Marketing Research* (Vol. 14, Issue 1). Yale University. <https://doi.org/10.2307/3151073>
- MURTI, M. A. W. K. (2017). Penerapan Metode *K-Means Clustering* Untuk Mengelompokkan Potensi Produksi Buah – Buahan Di Provinsi Daerah Istimewa Yogyakarta. In Skripsi. Sanata Dharma University.
- Muttaqin, M. R., & Defriani, M. (2020). Algoritma *K-Means* untuk Pengelompokan Topik Skripsi Mahasiswa. *ILKOM Jurnal Ilmiah*, 12(2), 121–129. <https://doi.org/10.33096/ilkom.v12i2.542.121-129>
- Pakpahan, H. S., Widiyans, J. A., Daffa, H., Firmanda, A., & Basani, Y. (2022). Implementasi Metode *K-Means* Untuk Pengelompokan Potensi Produksi Komoditas Perkebunan. *Adopsi Teknologi Dan Sistem Informasi (ATASI)*, 1(1), 52–60. <http://e-journals2.unmul.ac.id/index.php/atasi/index>
- Ridzuan, F., & Wan Zainon, W. M. N. (2019). A review on *data cleansing methods for big data*. *Procedia Computer Science*, 161, 731–738. <https://doi.org/10.1016/j.procs.2019.11.177>
- Rustam, S. (2018). Analisa *Clustering Phising* Dengan *K-Means* Dalam Meningkatkan Keamanan Komputer. *ILKOM Jurnal Ilmiah*, 10(2), 175–181. <https://doi.org/10.33096/ilkom.v10i2.309.175-181>
- Satria, W., Kusuma, P. D., & Irawan, B. (2019). Pengelompokkan Data Mineral Di Indonesia Menggunakan Metode *K-means Clustering*. *EProceedings ...*, 6(2), 5674–5681. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10678>
- Siyamto, Y. (2017). Pemanfaatan *Data mining* Dengan *Metode Clustering* Untuk Evaluasi Biaya Dokumen Ekspor di PT Winstar Batam. *Media Informatika Budidarma*, 1(2), 28–31.
- Sofiyan, A., & Azkiya, A. (2022). Penerapan Metode Rough Set Menganalisis Penyakit Yang Sering Dikeluhkan Pasien (Studi Kasus Puskesmas Jaya Mukti Dumai). *INFORMATIKA*, 14(1), 31. <https://doi.org/10.36723/juri.v14i1.348>