

INVENTORI ANALITIK: PENENTUAN JUMLAH KATEGORISASI RAW MATERIAL PALING OPTIMAL DENGAN MENGGABUNGAN METODE WEIGHTED K-MEANS DAN FUZZY C MEANS (ALGORITMA PEMBELAJARAN TANPA PENGAWASAN) PADA DATA MULTIDIMENSI KUALITAS BATUBARA

¹Yoga P. Adiwiguna ²Mohamad N. Saleh ³Hadi Syuhara ⁴Hario Purbaseno ⁵Almira D. Kusuma

^{1,2,3,4,5}Adaro Indonesia

Artikel masuk : 11-04-2022 , Artikel diterima : 29-08-2022

Kata kunci: K-Means, Fuzzy C Means, Pembelajaran Tanpa Pengawasan, Kualitas Batubara, Inventori Analitik

Keywords: K-Means, Fuzzy C-Means, Unsupervised Learning, Coal Quality, Inventory Analytics

Abstrak

Memahami pentingnya peran analitik dalam pengelolaan inventori akan membantu perusahaan meningkatkan wawasan pada aktivitas inventarisasinya yang dapat menjelaskan kekuatan dan kelemahan paling krusial. Pendekatan ini memungkinkan bisnis untuk meningkatkan kendali produktivitas, efisiensi, dan profitabilitas terhadap setiap aktivitas pengelolaan inventory. Mengelola *raw material* sebagai salah satu aset inventori melibatkan beberapa aspek berwujud seperti (biaya modal, biaya penyimpanan tetap dan tidak tetap), dengan demikian usaha-usaha seperti penentuan berapa jumlah kategorisasi *raw material* yang harus dibentuk akan mempengaruhi fungsi operasional dan keuangan. Algoritma pengelompokan seperti K-Means dan Fuzzy C Means merupakan teknik yang digemari dalam analisis data statistik dan banyak digunakan dalam bidang pembelajaran mesin dan pengenalan pola. Memanfaatkan teknik ini untuk kategorisasi kualitas batubara sebelum di blending di *Run of Mine* sebagai *raw material* akan memberikan perspektif baru pembagian *interval* nilai kualitas batubara. Peran algoritma ini utamanya adalah membagi populasi data menjadi beberapa kelompok, sehingga data yang memiliki karakteristik yang sama akan disegmentasikan pada kelompok yang sama sedangkan data yang berada di luar karakteristik tersebut akan disegmentasikan pada kelompok data lainnya. Parameter ujinya berupa total variabilitas terkecil (WCSS), kohesi terbesar (*Silhouette Coefficient*), dan koefisien partisi terbesar. Menggabungkan kedua algoritma ini dan indeks validasinya akan memberikan rekomendasi jumlah kategorisasi kualitas batubara paling efisien dengan pertimbangan parameter paling optimal di setiap tumpukan. Sehingga pada tahap operasional dengan kuantitas yang sama, kategori tumpukan yang baru akan membantu menghemat ruang inventori dan tetap mampu memberikan performa variasi yang tidak jauh berbeda.

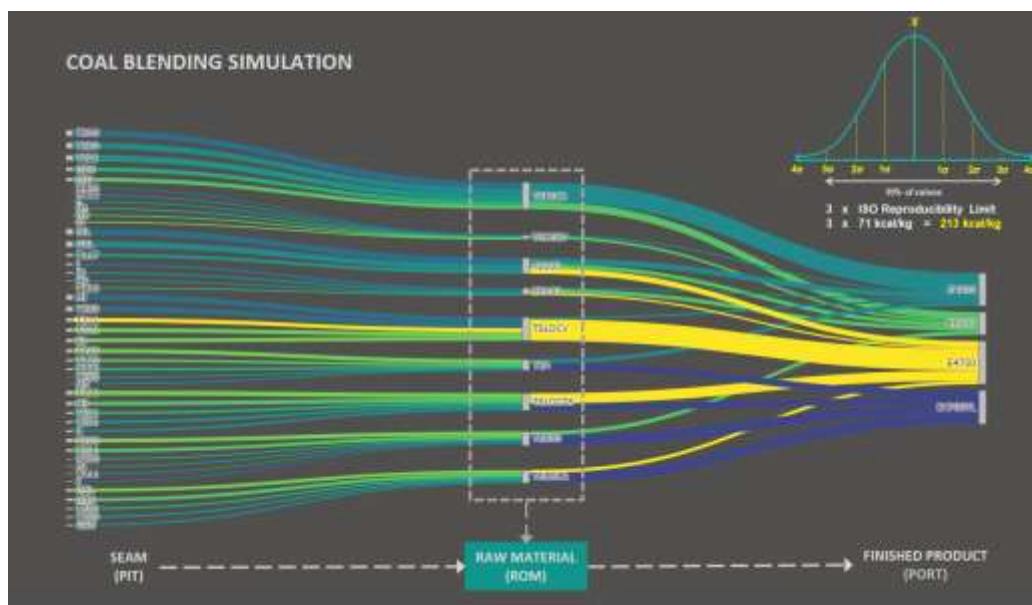
Abstract

Understanding the important role of inventory analytics will help companies gain insight into their inventory activity that defines crucial strengths and weaknesses. This approach enables businesses to improve their inventory control, productivity, efficiency, and profitability. Managing raw materials as a part of the inventory asset involves several tangible aspects (capital, carrying, and storage space cost), thus the efforts such as determining how many raw material categorizations that have to be constructed will affect both operational and financial purposes. Clustering algorithms such as K-Means and Fuzzy C Means are favored techniques for statistical data analysis used in many fields including machine learning and pattern recognition. Utilizing these techniques on categorizing coal quality before being blended into Run of Mine as raw materials will reveal a new perspective on dividing each interval of coal quality value. Their tasks mainly are to divide the data population into several groups, so data points which belong to the same characteristics will be segmented into the same group while data points that are outside these characteristics will be segmented into different groups. The test parameters are considering the smallest total variability (WCSS), the largest cohesion (Silhouette Coefficient), and the largest partition coefficient. Combining these two algorithms and their validation indexes will provide

recommendations for the most efficient number of coal quality categorizations by considering the local optima parameters in each coal stockpile. So later at the operational stage, the newly generated stockpile category would help save inventory space with the same quantity and still be able to provide variation performance at the same level.

PENDAHULUAN

PT. Adaro Indonesia menerapkan metode penumpukan sementara batubara di *Run of Mine* setelah proses penambangan sebelum dilakukan pengiriman ke *processing plant*. Tahapan ini dilakukan dengan tujuan memastikan kembali kondisi batubara sudah memenuhi standar kualitas yang telah ditetapkan. Kategorisasi batubara yang dilakukan saat ini mengacu pada nilai *interval* parameter *calorific value* 200 kcal/kg dan diberi penamaan khusus/labeling di masing-masing tumpukan (Gambar 1). Sehingga jika source batubara in situ memiliki kisaran *interval calorific value* 3800 – 5600 kcal/kg, maka nantinya di *Run of Mine* akan dilakukan kategorisasi batubara ke dalam 9 tumpukan berbeda seperti terlihat pada Gambar 2. Nilai kisaran *interval* 200 kcal/kg ini diterapkan karena mempertimbangkan 3σ level *reproducibility limit* atau masih mencakup 99,7 % probabilitas distribusi total pada nilai standar deviasi nya.



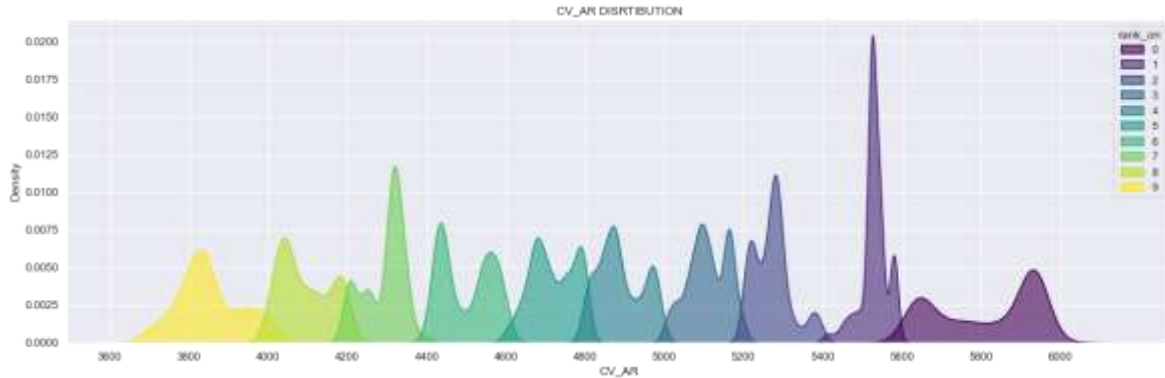
Gambar 1. Ilustrasi kategorisasi beberapa seam yang menjadi raw material

Penerapan metode ini terkadang menemui beberapa kendala pada tahap operasional, di antaranya:

1. Setiap tumpukan sering ditemui tidak memiliki pembagian kuantitas yang

seimbang, sehingga ada tumpukan yang memiliki cakupan area lebih besar dibandingkan tumpukan lainnya

2. Constraint operasional (*maintenance, unplanned stock*) menurunkan efisiensi space ROM dan menghambat FIFO.



Gambar 2. Interval calorific value 200 kcal dan density estimation pada raw material.

Kendala di atas merupakan hal yang dapat dirasakan langsung pada tahap operasional sedangkan dari sisi finansial akan mengakibatkan beban tambahan karena harus mempersiapkan area atau space yang lebih besar untuk mengakomodir banyaknya jumlah kategorisasi tumpukan, kemudian banyaknya jumlah unit yang harus disediakan untuk menunjang kesiapan area sekitar tumpukan dan tumpukan itu sendiri, terakhir adalah biaya keusangan yang harus dikeluarkan saat tumpukan tertahan cukup lama di *Run of Mine*. Kendala-kendala ini harus segera diselesaikan karena sangat berpengaruh terhadap fungsi operasional dan finansial. Beberapa alternatif solusi telah dilakukan di tahap operasional tetapi harus ada solusi yang tepat sasaran agar kemudian hari tidak muncul lagi aktivitas tidak terencana yang membutuhkan biaya tambahan. Sehingga dibutuhkan solusi berupa

penyederhanaan jumlah kategorisasi *raw material* tanpa mengorbankan performa akurasi yang ada.

METODOLOGI PENELITIAN

Penyederhanaan kategorisasi *raw material* melibatkan data proyeksi jangka panjang kualitas dan kuantitas batubara agar strategi yang nantinya diterapkan untuk mengelola inventori juga dapat berlangsung untuk periode yang cukup lama. Metode kategorisasi *raw material* yang diterapkan saat ini utamanya mempertimbangkan parameter *Calorific Value*, diikuti parameter lain seperti *Ash Content*, dan *Total Sulfur*, tetapi dalam penelitian ini penulis mencoba melibatkan parameter tambahan lainnya seperti *Inherent Moisture*, *Fixed Carbon*, dan *Volatile Matter* (Tabel 1)

Tabel 1. Sebagian data yang digunakan untuk mendukung penelitian

TAHUN	INTERVAL	VOLUME (TON)	ASH (ADB)	FC (ADB)	IM (ADB)	TS (ADB)	VM (ADB)	CV (AR)
2021	T200	505.807	2,2	36,5	19,7	0,09	41,6	4778
2021	T210	589.770	2,1	36,5	20,1	0,09	41,3	4781
2021	T300	611.692	2,1	36,5	20,3	0,09	41,2	4782
2021	T200	569.219	2,1	36,5	20,3	0,09	41,2	4784
2022	T220	504.310	1,2	37,3	19,7	0,09	41,8	4911
2022	T200	520.509	1,2	37,3	19,7	0,09	41,8	4919
2041	T222L	511.729	5,8	35,3	22,4	0,34	37,7	3975
2041	T310	688.289	3,5	34,0	25,3	0,21	38,4	4100
2042	T311L	700.627	3,1	35,2	22,8	0,16	39,6	4193
2042	T311	713.072	3,1	35,4	22,4	0,16	39,7	4203

Agar dapat melihat sebaran *datapoint* dari tabel di atas, menggunakan parameter utama

calorific value akan coba diplotkan ke dalam grafik 1D seperti terlihat pada Gambar 3.



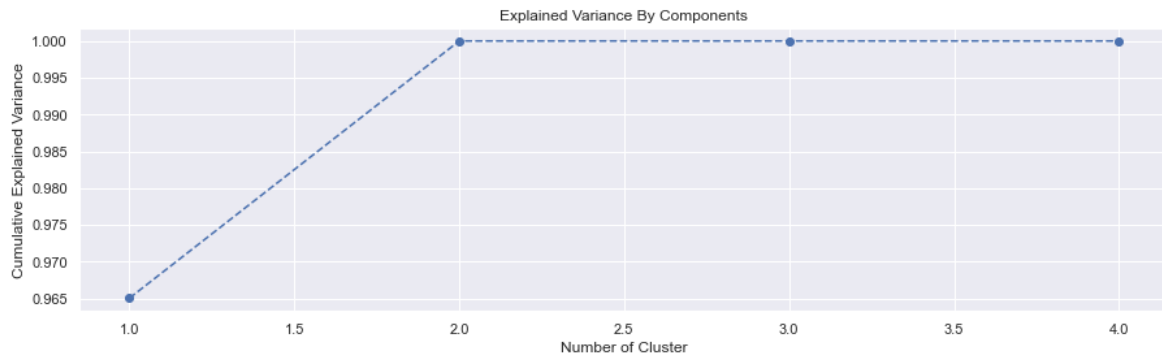
Gambar 3. Distribusi *datapoint* parameter *calorific value* dalam grafik 1D

Terlihat sekilas bahwa sebaran *data point* pada grafik 1D membentuk beberapa kelompok data yang masih dapat diidentifikasi, tetapi pada *interval* nilai *calorific value* 4500 – 5400 sebaran datanya sangat sulit diidentifikasi jika terdapat kelompok data baru karena rapatnya jarak antar *datapoint* pada *interval* tersebut. Sehingga dibutuhkan visualisasi pada dimensi yang lebih tinggi agar diperoleh gambaran sebaran data yang lebih baik. Untuk menghasilkan grafik 2D harus melibatkan minimal 2 parameter sebagai input, tetapi dengan banyaknya parameter yang dimiliki tidak dimungkinkan untuk mengikutsertakan

keseluruhan parameter tanpa menggunakan teknik reduksi dimensi. Sehingga 5 parameter pendukung (IM, VM, FC, Ash, TS) akan dilebur dan direduksi menjadi parameter baru (parameter baru < jumlah parameter awal) tetapi masih mewakili sebagian besar informasi yang terdapat dalam dataset sebelumnya menggunakan teknik *Principal Component Analysis* (Persamaan 1). Akan menjadi berapa *cluster* reduksi yang harus dipilih harus memenuhi minimal 80% *Cumulative Explained Variance* (Persamaan 2) seperti terlihat juga pada Gambar 4.

$$Final\ Dataset\ PCA = FeatureVector^T \times StandardizedOriginalDataset^T \quad (1)$$

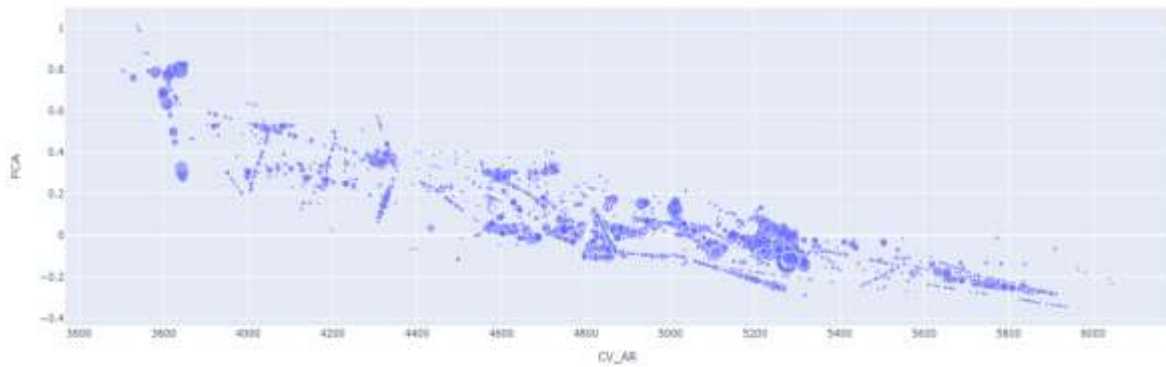
$$Explained\ Variance\ by\ component = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \dots + \lambda_n} \quad (2)$$



Gambar 4. *Cumulative Explained Variance*

Maka dari grafik di atas 5 parameter awal bisa diwakilkan ke dalam 1 parameter baru karena nilai *Cumulative Explained Variance* dengan

hanya 1 parameter baru bisa mewakili 96,5% informasi yang terdapat pada dataset dengan 5 parameter sebelumnya.



Gambar 5. Distribusi data point dalam grafik 2D

Dari parameter perwakilan baru tersebut akan coba dibandingkan dengan parameter utama *calorific value* pada grafik 2D dan diperoleh gambaran distribusi *datapoint* yang lebih baik (Gambar 5). Jika ingin membagi *datapoint* ini menjadi *cluster-cluster* tersendiri dengan hanya bantuan kasat mata pasti akan menemui kesulitan, sehingga dibutuhkan bantuan komputasi dengan teknik pembelajaran mesin

(*Machine Learning*) agar mempermudah mengidentifikasi *cluster* dibantu dengan indeks validasinya. Beberapa algoritma yang sering digunakan untuk tugas *clustering* dan segmentasi di antaranya adalah K-Means, DBSCAN, PAM, CLARA, sedangkan pemilihan algoritma mana yang sesuai dengan studi kasus dan dataset yang dimiliki mengacu pada Tabel 2 dan Tabel 3.

Tabel 2. Survei komprehensif pada beberapa algoritma *clustering* (Xu D, 2015)

CATEGORY	TYPICAL ALGORITHM	COMPLEXITY (Time)	SCALABILITY	FOR LARGE-SCALE DATA	FOR HIGH-DIMENSIONAL DATA	SHAPE OF SUITABLE DATASET	SENSITIVE TO THE SEQUENCE OF INPUTTING DATA	SENSITIVE TO NOISE/OUTLIER
Based on Partition	K-Means	Low $O(kn)$	Middle	Yes	No	Convex	Highly	Highly
	K-Medoids	High $O(k(n-k)^2)$	Low	No	No	Convex	Moderately	Little
	PAM	High $O(k^2n^2)$	Low	No	No	Convex	Moderately	Little
	CLARA	Middle $O(kn^2 + k(n-k))$	High	Yes	No	Convex	Moderately	Little
	CLARANS	High $O(n^2)$	Middle	Yes	No	Convex	Highly	Little
Based on Hierarchy	AP	*	*	*	*	*	*	*
	BIRCH	Low $O(n)$	High	Yes	No	Convex	Moderately	Little
	CURE	Low $O(s^2 \log n)$	High	Yes	Yes	Arbitrary	Moderately	Little
	ROCK	High $O(n^2 \log n)$	Middle	No	Yes	Arbitrary	Moderately	Little
	CHAMELEON	High $O(n^2)$	High	No	No	Arbitrary	Moderately	Little
Based on Fuzzy Theory	FCM	Low $O(n)$	Middle	No	No	Convex	Moderately	Highly
	FCS	High (kernel)	Low	No	No	Arbitrary	Moderately	Highly
	MM	Middle $O(n^2)$	Low	No	No	Arbitrary	Moderately	Little

Tabel 3. Perbandingan fitur-fitur pada beberapa algoritma *clustering* (Chitra K, 2017)

ALGORITHM	SCALABILITY & EFFICIENCY	NOISE	SHAPE OF CLUSTER	INPUT DATA
K-Means	Scalable in processing large datasets	Sensitive to noise and outliers	Works well only with clusters of convex shapes	Works only on numerical data
PAM [3]	Works well for small datasets but not for large datasets	Not very sensitive to noise and outliers		Works on data of all attributes
CLARA [3]	Can deal with larger datasets in comparison to PAM. Efficiency depends on sample size	Not very sensitive to noise and outliers		Works on data of all attributes
BIRCH	One of the best algorithms for large databases in terms of running time, space required, quality, number of I/O operations applied. Shows linear scalability with respect to a number of objects		Performs clustering well only with spherical data	Works on data of all attributes
CHAMELEON			Good at finding clusters of arbitrary shape	Works on data of all attributes
DBSCAN	Does not work well for high dimensional data	Handles noise effectively	Good at finding clusters of arbitrary shape	
DENCLUE	Does not work well for high dimensional data	Invariant against noise	Can find clusters of arbitrary shape	
STING				Used mainly with numerical values
CLIQUE	Scales linearly with the size of the input and shows good scalability when the number of dimensions are increased			It is not sensitive to input order

Sehingga dipilih algoritma K-Means dan Fuzzy C Means karena beberapa pertimbangan seperti waktu pemrosesannya yang cepat, skalabilitas menengah artinya waktu pemrosesan bisa lebih dipercepat seiring dengan ditingkatkannya jumlah dan performa mesinnya, juga cukup sensitif terhadap *noise* dan *outliers*. Kedua algoritma ini juga memiliki pendekatan yang hampir sama dalam penentuan kelompok data karena setiap observasinya mempertimbangkan pencarian pusat *cluster* atau *centroid cluster* terdekat, hanya saja algoritma FCM mengarah pada *soft clustering* atau memungkinkan setiap *datapoint* masuk ke lebih dari satu *cluster*. Indeks validasi yang digunakan sebagai penentuan baik tidaknya jumlah *cluster* yang harus dipilih berupa *Within Cluster Sum of Square* terkecil (Persamaan 3), *Silhouette*

Coefficient terbesar (Persamaan 4), dan *Partition Coefficient* terbesar (Persamaan 5)

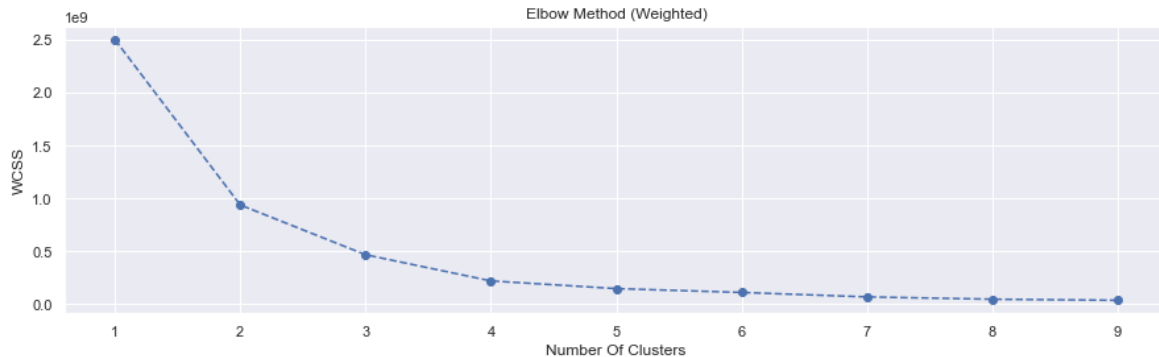
$$WCSS = \sum_{i=1}^k \sum_{x \in S_i} ||x - \mu_i||^2$$

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}}, \text{ if } |C_i| > 1$$

$$PC = \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^N u_{ci}^2 \quad (5)$$

Setelah diperoleh rekomendasi jumlah *cluster* terbaik, akan dilakukan visualisasi ulang pada dataset yang sudah memiliki label asosiasi untuk masing-masing *cluster* baru, kemudian dilakukan evaluasi distribusi volume mengacu pada parameter *calorific value* yang sebelumnya menggunakan *interval* 200 kcal/kg dibandingkan dengan *interval calorific value* baru yang dibentuk oleh hasil algoritma *clustering*.

HASIL DAN PEMBAHASAN



Gambar 6. Nilai WCSS pada masing-masing *cluster*

Dari nilai WCSS (Gambar 6) terlihat bahwa jumlah variabilitas *inter cluster* mulai mengecil signifikan saat berada di slope garis ke 3 dan 4, indeks ini mengindikasikan

pemilihan *cluster* sudah bisa dilakukan mulai *cluster* ke 3 dan 4, karena dengan pemilihan jumlah *cluster* yang lebih banyak tidak akan menurunkan variabilitas secara signifikan lagi.

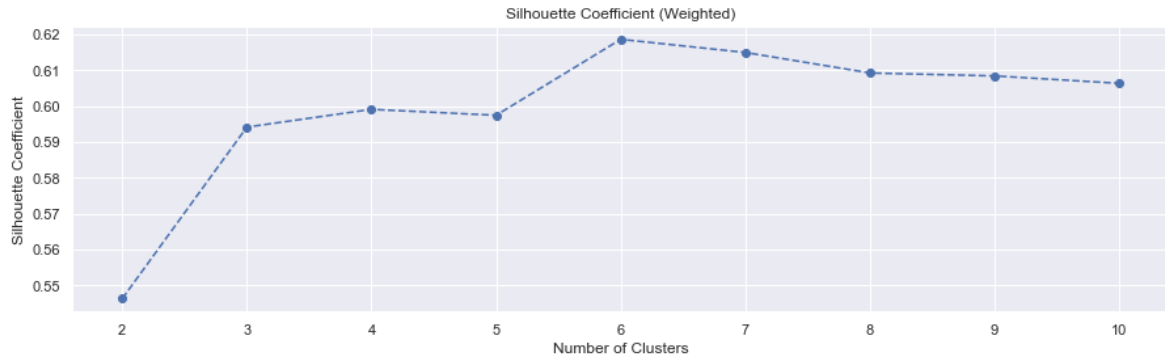
```
In [17]: from kneed import KneeLocator
kl = KneeLocator(range(1, 10), wcss, curve="convex", direction="decreasing")
kl.elbow

Out[17]: 3
```

Gambar 7. Penentuan kurvatur maksimum

Agar tidak bias dalam pemilihan *slope* awal untuk penentuan *cluster* di kurva WCSS, digunakan teknik *knee locator* dan dihasilkan

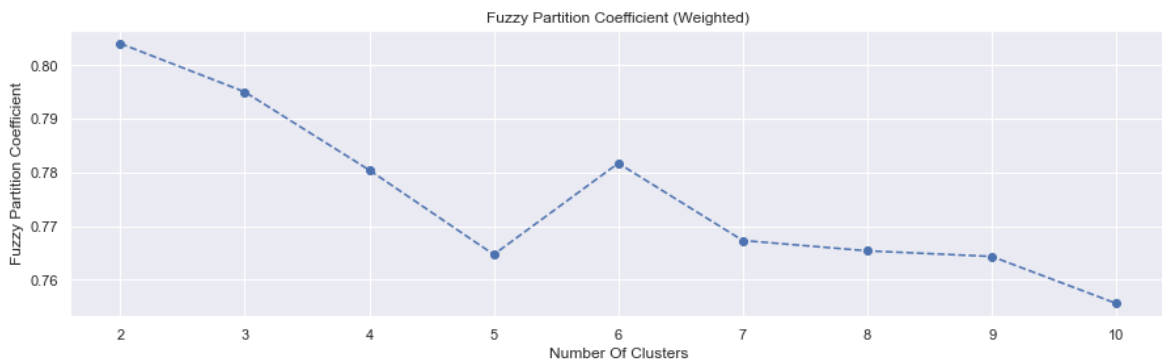
bahwa kurvatur maksimum berada di titik ke 3 (Gambar 7).



Gambar 8. Nilai *Silhouette Coefficient* pada masing-masing *cluster*

Kemudian untuk nilai *Silhouette Coefficient* (Gambar 8) diperoleh nilai terbesar derajat keasaman atau konsistensi *inter cluster*

(kohesi) dibandingkan dengan *cluster* lain (separasi) saat data dibagi menjadi 6 *cluster*.



Gambar 9. Nilai Fuzzy Partition Coefficient pada masing-masing *cluster*

Menggunakan matriks validasi FPC (Gambar 9), secara berurutan diperoleh 3 nilai FPC

tertinggi saat data dikelompokkan menjadi 2 *cluster*, 3 *cluster*, dan 6 *cluster*.

Tabel 4. Perbandingan nilai matriks validasi pada setiap jumlah *cluster*

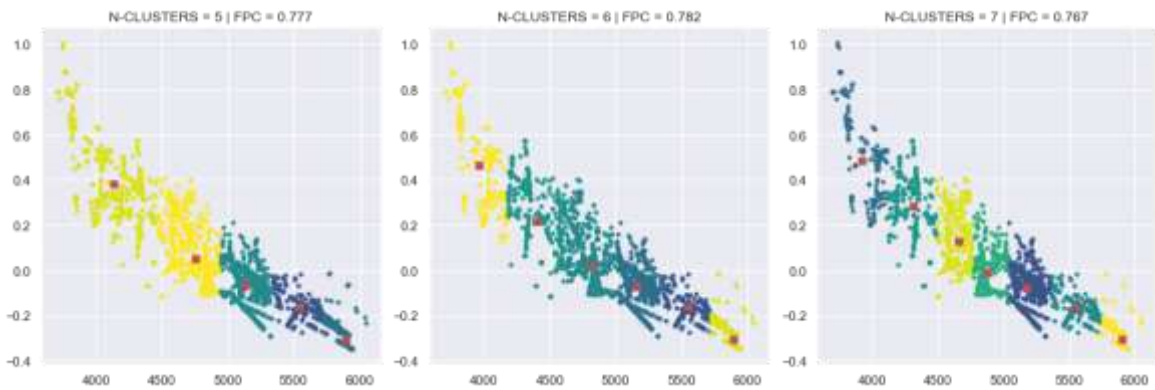
MATRIKS	C.2	C.3	C.4	C.5	C.6	C.7
WCSS	0.463×10^{10}	0.208×10^{10}	0.105×10^{10}	0.073×10^{10}	0.047×10^{10}	0.033×10^{10}
Silhouette Coef.	0.546	0.594	0.599	0.600	0.618	0.614
FPC	0.804	0.795	0.780	0.765	0.782	0.767

Dari ketiga matriks validasi yang dihasilkan (WCSS, *Silhouette Coefficient*, FPC) seperti terlihat pada Tabel 4, mengindikasikan bahwa pembagian data sudah bisa dilakukan mulai dengan 3 *cluster* hasil dari nilai kurvatur maksimum WCSS dan nilai FPC tertinggi kedua dihasilkan pada saat data dibagi ke dalam 3 *cluster*, tetapi jumlah *cluster* sebanyak ini sebenarnya akan menghasilkan performa variasi parameter *calorific value* yang tidak baik, karena *interval* yang dihasilkan saat sudah menjadi *raw material* pasti akan sangat lebar (± 1.000 kcal/kg) hal ini bisa dilihat dari

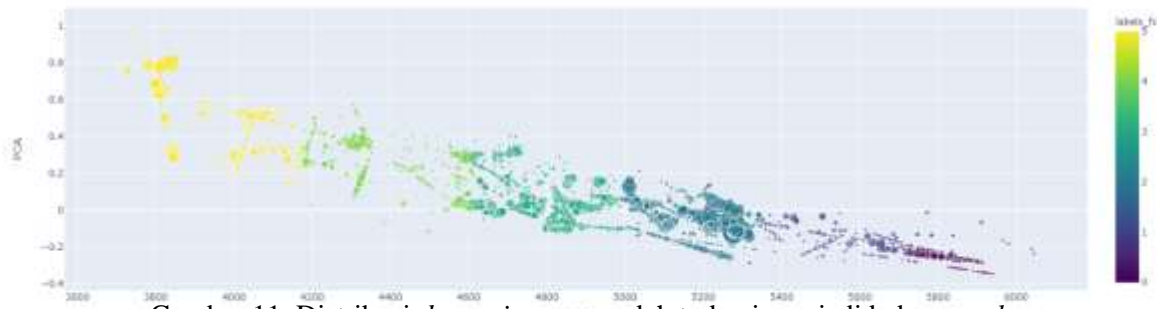
jarak *centroid* pada Gambar 10. Dengan adanya pertimbangan tersebut maka dipilih nilai perbandingan berikutnya di mana nilai perbandingan tertinggi dihasilkan pada saat data dibagi ke dalam 6 *cluster*, FPC tertinggi setelah *cluster* ke 2 dan ke 3 dibentuk saat data juga dibagi ke dalam 6 *cluster*, sedangkan untuk nilai WCSS malah akan menghasilkan variasi *inter cluster* yang semakin kecil jika semakin besar jumlah *cluster* yang dipilih. Dari keseluruhan pertimbangan tersebut maka jumlah optimal kategorisasi data seam batubara sebanyak 6

cluster atau 6 jenis raw material (Gambar 11) yang akan menjadi acuan baru penumpukan di Run of Mine. Nilai density estimator dan

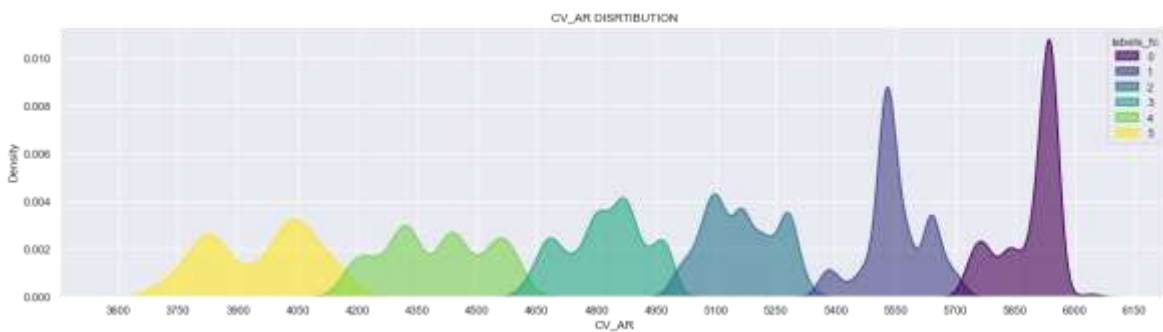
interval yang dibentuk oleh algoritma ini berurutan terlihat pada Gambar 12 dan Tabel 5.



Gambar 10. Lokasi centroid yang dibentuk pada beragam jumlah cluster



Gambar 11. Distribusi datapoint yang sudah terbagi menjadi beberapa cluster



Gambar 12. Interval calorific value dan density estimation baru hasil algoritma clustering

Tabel 5. Interval calorific value baru yang dibentuk oleh algoritma clustering

Cluster	Min	Max	Diff.
0	5725	6045	320
1	5353	5722	368
2	4987	5350	363
3	4609	4986	377
4	4178	4607	429
5	3704	4178	474

KESIMPULAN

Beberapa poin yang dapat disimpulkan dari penelitian ini antara lain:

1. Jumlah rekomendasi optimal *cluster* seam menjadi *raw material* berubah dari awalnya 9 kategori menjadi hanya 6 kategori
2. Nilai *interval raw material* baru pada parameter *calorific value* yang dibentuk pada masing-masing *cluster* bervariasi dari 320 kcal/kg hingga 470 kcal/kg
3. Walaupun nilai *interval* baru yang dibentuk oleh algoritma *clustering* lebih lebar dibandingkan *interval* metode kategorisasi per 200 kcal/kg, tetapi secara *density estimation interval* baru ini lebih stabil karena dominan membentuk kurva ideal distribusi normal
4. Jika ingin mempertahankan jumlah kategorisasi *raw material* yang ada saat ini, bisa mempertimbangkan fleksibilitas nilai *interval* masing-masing *raw material* menggunakan algoritma *clustering* yang sama, agar dicapai nilai *interval calorific value* yang lebih kecil tetapi *density estimation* yang lebih baik dari pada *interval* 200 kcal/kg
5. Implementasi algoritma *clustering* bisa dilakukan untuk penentuan kategorisasi *raw material* periode data *reserve* batubara jangka pendek atau jangka panjang menyesuaikan strategi penumpukan dan administrasi di lapangan

UCAPAN TERIMA KASIH

Pada kesempatan ini, kami sampaikan terima kasih sebesar-besarnya kepada PERHAPI karena telah menyelenggarakan TPT XXX PERHAPI 2021. Acara ini sangat mendukung pengembangan ilmu pengetahuan dan memacu kemunculan terobosan baru di bidang pertambangan. Tidak lupa juga kami ucapkan terima kasih kepada seluruh pihak yang telah mendukung penulisan hingga penerbitan makalah ini, semoga Tuhan Yang Maha Esa menganugerahkan balasan yang setimpal atas kebaikan Bapak/Ibu sekalian.

DAFTAR PUSTAKA

- Chitra, K. (2017): A Comparative Study of Various Clustering Algorithms in Data Mining, *International Journal of Computer Science and Mobile Computing*, **6**, 109 – 115.
- Feature Engineering*, data diperoleh melalui situs internet: <https://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>. Diakses pada tanggal 9 Agustus 2021
- Feature Weighted Fuzzy C Means Classifier*, data diperoleh melalui situs internet: https://github.com/Csinclair0/feature_learning/blob/master/Notebooks/FeatureWeightedFuzzyX.ipynb. Diakses pada tanggal 9 Agustus 2021
- Fuzzy Clustering*, data diperoleh melalui situs internet: https://en.wikipedia.org/wiki/Fuzzy_clustering#Fuzzy_c-means_clustering. Diakses pada tanggal 9 Agustus 2021
- Fuzzy Logic Algorithm*, data diperoleh melalui situs internet: https://pythonhosted.org/scikitfuzzy/aut_o_examples/plot_cmeans.html. Diakses pada tanggal 9 Agustus 2021.
- International Organization for Standardization. (2009). *Solid mineral fuels – Determination of gross calorific value by bomb calorimetric method and calculation of net calorific value* (ISO Standard No. 1928:2009).
- Harrington, P. (2012): Grouping Unlabeled Items, *Machine Learning in Action*, Manning Publications, Shelter Island, New York, 208 – 209.
- Ivanov, D. (2019): Role, Functions, and Types of Inventory, *Global Supply Chain and Operations Management*, Springer Nature Switzerland, Cham, Switzerland, 363 – 365.
- K-Means Clustering*, data diperoleh melalui situs internet: https://en.wikipedia.org/wiki/Kmeans_clustering. Diakses pada tanggal 9 Agustus 2021
- K-Means Clustering in Python*, data diperoleh melalui situs internet: <https://realpython.com/kmeansclustering-python/>. Diakses pada tanggal 9 Agustus 2021.

- King, R.S. (2014): Fuzzy Neighborhood, *Cluster Analysis and Data Mining an Introduction*, Mercury Learning and Information, Dulles, Virginia, United States of America, 119.
- Knee Locator*, data diperoleh melalui situs internet:
<https://kneed.readthedocs.io/en/stable/parameters.html>. Diakses pada tanggal 9 Agustus 2021
- Principal Component Analysis in Python*, data diperoleh melalui situs internet:
<https://365datascience.com/tutorials/python-tutorials/pca-k-means/>. Diakses pada tanggal 9 Agustus 2021
- Speight, J.G. (2015): Thermal properties, *Handbook of Coal Analysis*, John Wiley & Sons Inc, Hoboken, NJ, USA, 198 – 225.
- Stadtler, H. (2005): Inventory Analysis, *Supply Chain Management and Advanced Planning*, Springer Berlin, Heidelberg, New York, 56 – 57.
- Thakur, A. (2020): Decomposition Techniques, *Approaching Almost Any Machine Learning Problem* [Kindle Version], Diambil dari situs internet <https://www.amazon.com/>
- Weighted Location Clustering*, data diperoleh melalui situs internet:
<https://github.com/JosephMagiya/WeightedLocationClustering/blob/master/Weighted%20Location%20Clustering.ipynb>. Diakses pada tanggal 9 Agustus 2021.
- Xu, D. (2015): A Comprehensive Survey of Clustering Algorithms, Springer-Verlag Berlin Heidelberg, 165 – 193.