



## Klasterisasi Mineral Batuan di Lapangan X berdasarkan Data Spektral menggunakan K-Means Clustering

Sulaiman Abdullah Pane\*, Felix Mulia Hasudungan Sihombing

Fakultas Matematika dan Ilmu Pengetahuan Alam, Program Studi Geologi, Universitas Indonesia, Depok, Indonesia

Email: <sup>1,\*</sup> [sulaiman.abdullah@ui.ac.id](mailto:sulaiman.abdullah@ui.ac.id), <sup>2</sup> [felix.sihombing@ui.ac.id](mailto:felix.sihombing@ui.ac.id)

Email Penulis Korespondensi: [Sulaiman.abdullah@ui.ac.id](mailto:Sulaiman.abdullah@ui.ac.id)

**Abstrak**—Teknologi terus diterapkan dalam bidang geologi pada berbagai cabang keilmuannya, salah satunya adalah pemanfaatan metode pembelajaran mesin yang termasuk dalam teknologi kecerdasan buatan. Identifikasi mineral batuan dapat dilakukan menggunakan metode pembelajaran mesin. Klasterisasi mineral batuan dilakukan untuk mengidentifikasi persebaran jumlah kelompok mineral teroptimal berdasarkan informasi geologi yang dimiliki pada data hasil pengeboran batuan dalam tahap eksplorasi geologi di wilayah Manjimup, Australia Bagian Barat. Identifikasi mineral batuan melalui klasterisasi dilakukan menggunakan pembelajaran mesin tidak terawasi (*unsupervised*) dengan metode *K-Means clustering*. Data yang digunakan dalam adalah data hasil pengukuran spektrum elektromagnetik berupa data spektral *Thermal Infrared (TIR)* yang berasal dari data hasil pengeboran batuan. Data spektral yang digunakan terdiri dari 341 parameter sehingga dilakukan pengurangan dimensi input untuk mengurangi kompleksitas komputasi menggunakan *Principal Component Analysis (PCA)* menjadi data dua dimensi agar dapat divisualisasikan dengan lebih mudah. Berdasarkan hasil evaluasi, jumlah kelompok mineral batuan teroptimal melalui hasil klasterisasi menggunakan *K-Means* berdasarkan informasi geologi yang dimiliki adalah 3 kelompok mineral batuan.

**Kata Kunci:** Data Spektral, Klasterisasi, *K-Means*, Mineral, Pembelajaran Mesin, *Principal Component Analysis (PCA)*

**Abstract**—Technology continues to be applied in the field of geology in various branches of science, one of which is the use of machine learning methods which are included in artificial intelligence technology. Machine learning methods able to identifying rock minerals. Rock mineral clustering is carried out to identify the distribution of the optimal number of mineral groups based on geological information held in rock drilling results data during the geological exploration stage in the Manjimup region, Western Australia. Identification of rock minerals through clustering is carried out using unsupervised machine learning with the *K-Means clustering* method. The data used in this research are data from the measurement of the electromagnetic spectrum in the form of *Thermal Infrared (TIR)* spectral data derived from rock drilling results. The spectral data used consisted of 341 parameters so that the input dimension was reduced to reduce computational complexity using *Principal Component Analysis (PCA)* into two-dimensional data so able to visualized more easily. Based on the evaluation results, the optimal number of rock mineral groups through the results of clustering using *K-Means* based on geological information is 3 groups of rock minerals.

**Keywords:** *Spectral Data, Clustering, K-Means, Minerals, Machine Learning, Principal Component Analysis (PCA)*

### 1. PENDAHULUAN

Perkembangan zaman yang terus diikuti dengan pemanfaatan teknologi pada era revolusi industri 4.0 saat ini telah banyak tercipta berbagai teknologi yang bermanfaat untuk mempermudah pekerjaan yang dilakukan oleh manusia dalam kehidupannya. Revolusi industri 4.0 telah mempengaruhi esensi dalam kehidupan manusia secara luas [1].

Salah satu pemanfaatan teknologi melalui revolusi industri 4.0 adalah teknologi kecerdasan buatan yang diterapkan dalam berbagai disiplin ilmu. Dalam bidang geologi, teknologi kecerdasan buatan telah digunakan dalam proses identifikasi batuan menggunakan pembelajaran mesin [2]. (Pembelajaran mesin telah dimanfaatkan dalam bidang geologi terutama pada sektor eksplorasi sumber daya energi untuk melakukan klasifikasi berbagai aspek geologi termasuk jenis batuan, jenis mineral, struktur geologi, serta aspek lainnya [3].

Pembelajaran mesin merupakan salah satu cabang ilmu dari teknologi kecerdasan buatan yang mampu membuat komputer melakukan serangkaian kegiatan pembelajaran secara mandiri tanpa disertai pemrograman secara eksplisit [4]. Dalam kegiatan pembelajaran mesin, terdapat beberapa metode yang dapat digunakan sesuai *output* yang ingin dihasilkan dan ketersediaan data yang dimiliki. Salah satunya adalah pembelajaran mesin tidak terawasi dimana pembelajaran dilakukan tanpa disertai label data pada data *input* sehingga mesin akan mempelajari informasi data input untuk menghasilkan pola tertentu [5]. *Output* yang dihasilkan melalui pembelajaran tersebut adalah klasterisasi atau pengelompokan data. Klasterisasi dapat dilakukan menggunakan metode *K-means* yang umum dilakukan pada berbagai jenis dalam kegiatan pengelompokan data menjadi sejumlah nilai K kelompok yang sejajar dengan kelompok lainnya sesuai dengan yang didefinisikan [6].

Penelitian terdahulu dari penelitian ini merupakan karakterisasi data *well log* berupa log densitas dan data kecepatan gelombang P menggunakan metode *k-means clustering* dimana dilakukan kegiatan klasterisasi untuk mengidentifikasi perbedaan jenis litologi batuan, pori fluida, dan kontak fluida [7]. Berbeda dengan penelitian terdahulu, penelitian ini dilakukan menggunakan pendekatan data spektral hasil pengukuran instrumen spektroskopi pada data hasil pengeboran batuan untuk menentukan persebaran kelompok mineral batuan berdasarkan metode *k-means clustering* yang belum pernah dilakukan sebelumnya.

Data yang digunakan dalam penelitian ini adalah data spektral berisikan informasi mengenai reflektansi pada panjang gelombang tertentu, kedalaman, kode sampel, dan jenis mineral yang berasal dari pengukuran



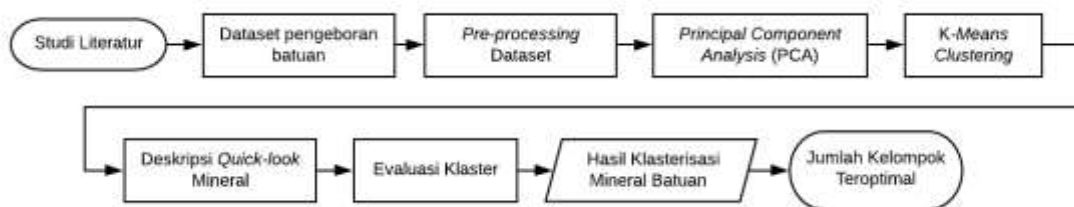
instrumen spektrum elektromagnetik. Jumlah parameter data yang mencapai 341 parameter membuat proses visualisasi data sulit sehingga dilakukan pengurangan dimensi yang lebih kecil agar dapat menghasilkan data dengan 2 dimensi menggunakan *principal component analysis* (PCA) [6].

Daerah penelitian terletak di wilayah Manjimup, Australia Bagian Barat yang memiliki geologi regional berupa deposit Au-Ag-Cu pada endapan *Volcanogenic Massive Sulphide* (VMS) [8]. Tujuan dilakukannya penelitian ini adalah untuk menentukan pengelompokan tipe batuan di lapangan X melalui metode pembelajaran mesin tidak terawasi pada data spektral *Thermal Infrared* (TIR) yang dikaitkan dengan informasi geologi yang terdapat pada daerah penelitian. Geologi spektral merupakan kegiatan yang melakukan pengukuran dan analisis bagian dari spektrum elektromagnetik untuk mengidentifikasi fitur-fitur batuan atau material yang terdapat dipermukaan secara fisik, mineralogi, dan tingkat alterasi dari batuan yang diidentifikasi [9].

## 2. METODOLOGI PENELITIAN

Metode penelitian dalam kegiatan penelitian ini terbagi menjadi beberapa tahapan yaitu tahap studi literatur, pengumpulan data hasil pengeboran batuan, reduksi dimensi data melalui metode PCA, klasterisasi menggunakan *K-means clustering*, deskripsi *quick-look* mineral, evaluasi klaster, lalu mendapatkan hasil klasterisasi mineral batuan dan memperoleh jumlah kelompok mineral batuan teroptimal. Berikut adalah rincian dari metode penelitian yang dilakukan:

- Studi literatur merupakan tahap persiapan dimana dilakukan pengumpulan dari berbagai literatur berupa buku, jurnal ilmiah, atau publikasi lainnya yang bersifat valid.
- Pengumpulan dataset pengeboran batuan merupakan tahap melakukan kegiatan pengambilan data untuk keperluan pengolahan data. Dataset terdiri dari 3 databor yaitu WPD02, WPD04, dan WPD07. Databor tersebut dalam penelitian ini digabung menjadi WPD247. Data yang digunakan dalam penelitian ini adalah data spektral. Data spektral dihasilkan melalui pengukuran oleh sensor spektral pada material (batuan inti) dengan prinsip sensor tersebut akan merekam radiasi yang dipantulkan dari permukaan material. Data tersebut mencakup informasi mengenai kode sampel batuan, kedalaman batuan, jenis mineral pada batuan, serta reflektansi mineral pada panjang gelombang tertentu. Panjang gelombang yang digunakan adalah *Thermal Infrared* (TIR).
- Pre-processing* dataset atau pemilahan dataset dilakukan untuk mengurangi kesalahan atau eror yang terjadi pada kegiatan pembelajaran mesin dengan cara membuang data yang tidak memiliki informasi (*Null*) atau bernilai 0.
- Analisis menggunakan *Principal Component Analysis* (PCA) untuk mengurangi dimensi pada dataset agar kompleksitas komputasi dataset berkurang.
- Pembelajaran mesin melalui bahasa pemrograman python menggunakan metode *K-Means clustering* untuk melakukan klasterisasi atau membagi kelompok mineral batuan berdasarkan nilai K atau jumlah kelompok yang ditentukan. Kemudian dilakukan penentuan nilai K optimal melalui metode *elbow*.
- Deskripsi *quick-look* mineral dilakukan setelah terbagi beberapa variasi jumlah kelompok mineral batuan yang belum memiliki informasi geologi didalamnya. Sehingga dilakukan deskripsi mineral untuk melengkapi informasi geologi.
- Evaluasi klaster dilakukan untuk menentukan jenis mineral batuan yang terkandung dalam setiap kelompok mineral sesuai dengan kondisi geologi dan hasil klasterisasi menggunakan metode *K-means*.
- Tahap hasil klasterisasi mineral batuan didapatkan kelompok mineral batuan yang teroptimal berdasarkan hasil pembelajaran mesin dan kondisi geologi daerah penelitian.



Gambar 1. Diagram Alir Penelitian

## 3. HASIL DAN PEMBAHASAN

### 3.1 Data Hasil Pengeboran Batuan

Data hasil pengeboran batuan yang akan diproses dalam kegiatan pembelajaran mesin berjumlah 61711 data yang telah dilakukan pemilahan data sehingga tidak terdapat data yang tidak memiliki informasi atau *null*. Data tersebut mencakup kode sampel, kedalaman sampel, jenis mineral, dan reflektansi pada panjang gelombang tertentu. Data



reflektansi memiliki interval panjang gelombang 25 nm dari rentang panjang 6000 – 14500nm sehingga memiliki 341 kolom data. Berikut merupakan tampilan dataset yang akan diproses.

```
In [2]: 1 df=pd.read_csv("C:\\Users\\baimp\\Desktop\\DATARAPIH\\TIR\\MINI\\247TIRMIN1CLUSTER.csv")
        2 df

out[2]:
```

	Sample	Depth	Mineral	WL1	WL2	WL3	WL4	WL5	WL6	WL7	...	WL332	WL333	WL
0	WPD02_0001_10	39.018222	Quartz	0.022803	0.024812	0.023224	0.024691	0.024372	0.024177	0.023824	...	0.021947	0.021190	0.02...
1	WPD02_0001_11	39.025737	Quartz	0.018675	0.019712	0.020718	0.021132	0.019932	0.017525	0.019339	...	0.018881	0.025890	0.03...
2	WPD02_0001_12	39.033243	Quartz	0.024539	0.021424	0.015185	0.020784	0.021417	0.020579	0.017868	...	0.006231	0.006092	0.00...
3	WPD02_0001_13	39.040768	Quartz	0.031015	0.028955	0.027794	0.025532	0.028409	0.027298	0.024491	...	0.019452	0.023895	0.02...
4	WPD02_0001_14	39.048264	Quartz	0.024064	0.027468	0.026317	0.024755	0.029027	0.024175	0.023963	...	0.021889	0.021932	0.02...
5	WPD02_0001_15	39.055779	Quartz	0.028667	0.024317	0.026355	0.022289	0.021471	0.020381	0.021432	...	0.029199	0.032030	0.03...
6	WPD02_0001_16	39.063285	Quartz	0.018141	0.021137	0.016955	0.015777	0.021522	0.021234	0.015199	...	-0.000904	0.001946	0.00...
7	WPD02_0001_17	39.070800	Gypsum	0.021376	0.021742	0.021373	0.018242	0.015521	0.014149	0.016159	...	0.007809	0.012559	0.01...
8	WPD02_0001_18	39.070808	Gypsum	0.037588	0.039421	0.037780	0.035438	0.031527	0.023174	0.024886	...	0.002483	0.002495	0.00...
9	WPD02_0001_19	39.078311	Gypsum	0.041188	0.040359	0.040990	0.038208	0.029782	0.025583	0.030567	...	0.008733	0.014684	0.02...
10	WPD02_0001_20	39.085827	Gypsum	0.034820	0.037002	0.034448	0.030374	0.029044	0.028486	0.027604	...	0.009510	0.011882	0.01...

**Gambar 2.** Tampilan kode memasukkan dataset pada Jupyter notebook

**3.2 Hasil Pembelajaran Mesin**

**3.2.1 Principal Component Analysis (PCA)**

Principal component analysis (PCA) dilakukan untuk mengurangi dimensi data. Data yang digunakan untuk melakukan metode tidak terawasi adalah data mineral TIR. Mineral *null* yang terdapat pada dataset tidak digunakan karena memiliki nilai kosong dan bukan termasuk dalam mineral. Hal tersebut dilakukan untuk menjaga kualitas data agar tidak terjadi kesalahan dalam proses pembelajaran mesin.. data yang terdiri dari 341 kolom akan direduksi dimensinya menjadi dua kolom atau dua dimensi yaitu sumbu X dan Y untuk memudahkan proses pembelajaran mesin [6]. Data yang telah direduksi oleh PCA menjadi dua dimensi tetap memiliki kualitas dan informasi yang sama dengan data asalnya. berikut merupakan tampilan data yang telah dilakukan PCA:

	principal component 1	principal component 2
0	-0.126687	-0.273041
1	0.038155	-0.294110
2	-0.485054	-0.435680
3	-0.036880	-0.313815
4	0.125660	-0.254127

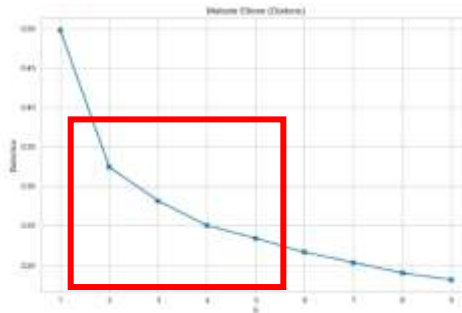
**Gambar 3.** Hasil PCA dataset

**3.2.2 K-Means Clustering**

Metode *k-means clustering* merupakan salah satu metode pembelajaran mesin tidak terawasi yang mampu melakukan klasterisasi atau pengelompokan data. Metode *k-means* dapat dilakukan secara cepat dan mudah untuk mengimplementasikannya sehingga digunakan dalam kegiatan klasterisasi [10]. Dalam *k-means*, k merupakan variabel angka merujuk pada jumlah klaster yang akan dibentuk. Metode *elbow* digunakan untuk menentukan jumlah k (klaster) yang paling optimal dengan berdasarkan nilai distorsi pada setiap nilai k sehingga membentuk garis siku [11] seperti pada gambar 4. Pada grafik tersebut, garis antara nilai k yang mengalami patahan akan membentuk siku (*elbow*) pada saat k = 2, k = 3, k = 4, k = 5. Terdapat 4 nilai yang diinterpretasikan sebagai nilai k paling optimal karena patahan yang dialami oleh data tidak begitu signifikan dalam menentukan 1 nilai terbaik sehingga metode *elbow* digunakan sebagai pertimbangan awal dalam menentukan klaster yang optimal berdasarkan bentuk siku dan nilai selisih distorsi yaitu pada saat nilai k = 2, k = 3, k = 4, k = 5. Penentuan jumlah nilai paling optimal lebih rinci dilakukan berdasarkan informasi geologi yang dicirikan dengan persebaran jenis mineral pada masing-masing kelompok hasil klasterisasi.

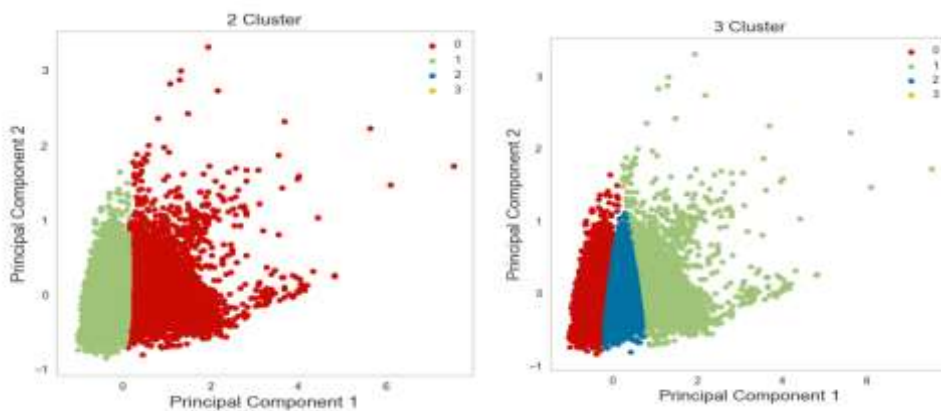
**Tabel 1.** Selisih nilai distorsi pada K

Nilai K	Nilai Distorsi	Selisih Nilai Distorsi
1	0.49	0.49
2	0.32	0.17
3	0.28	0.04
4	0.25	0.03
5	0.23	0.02
6	0.21	0.02
7	0.2	0.01
8	0.19	0.01

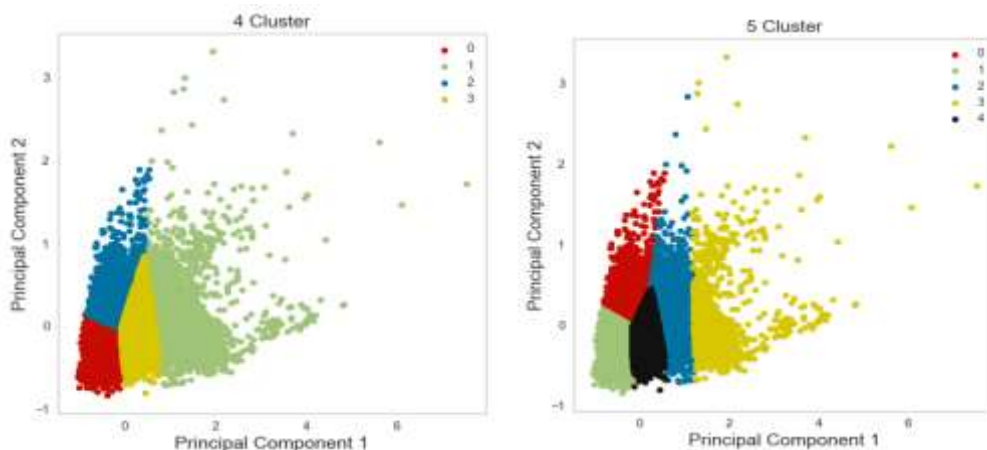


**Gambar 4.** Metode *elbow* pada dataset

Setelah melakukan analisis menggunakan metode *elbow* untuk menentukan pertimbangan awal dalam memilih jumlah kluster yang paling optimal, selanjutnya melakukan klusterisasi menggunakan metode k-means. Persebaran pada jumlah kluster 2 hingga 5 memiliki distribusi data kluster yang bervariasi. Pada jumlah 2 kluster, kluster 0 yang ditandai dengan warna merah memiliki persebaran distribusi yang lebih dominan. Pada jumlah 3 kluster, semua kluster memiliki distribusi yang mirip namun didominasi oleh kluster 1. Pada jumlah 4 kluster, terdapat perbedaan distribusi yang cukup signifikan dimana kluster 0 dan kluster 2 memiliki distribusi yang serupa dan kluster 3 mendominasi persebaran data. Pada jumlah kluster 5, terdapat kemiripan dalam distribusi persebaran data dengan 4 kluster dimana pada jumlah 5 kluster, kluster 2,3, dan 4 terbagi menjadi 3 kluster sedangkan pada jumlah 4 kluster terbagi menjadi 2 kluster.



**Gambar 5.** Persebaran hasil klusterisasi dengan jumlah (kiri: k=2, kanan: k=3)



**Gambar 6.** Persebaran hasil klusterisasi dengan jumlah (kiri: k=4, kanan: k=5)

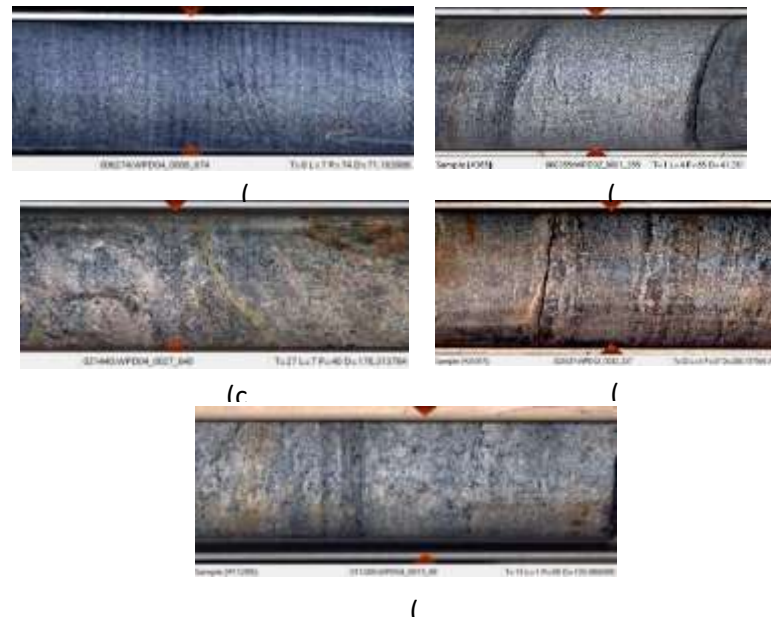
### 3.3 Deskripsi *Quick-look* Mineral

Deskripsi mineral batuan dilakukan secara manual menggunakan *linescan image* yang terdapat pada *software The Spectral Geologist (TSG)*. Deskripsi mineral batuan dilakukan untuk menentukan persebaran mineral pada dataset. Informasi mengenai deskripsi mineral batuan kemudian digunakan untuk melengkapi informasi kluster. Jumlah kluster yang telah ditentukan berdasarkan metode *elbow* belum memuat informasi mengenai mineral batuan yang terdapat dalam kluster tersebut. sehingga dengan melakukan deskripsi mineral batuan serta melakukan





perbandingan dengan data litologi batuan untuk setiap klaster, dapat menentukan jumlah klaster yang paling optimal berdasarkan informasi geologi. Deskripsi mineral dilakukan menggunakan acuan warna mineral dominan yang terdapat pada *linescan* batuan [12]. Hal tersebut dikarenakan ciri mineral yang dapat dideskripsi secara umum menggunakan *linescan image* secara *quicklook* adalah hanya warna mineral. Pembagian mineral secara garis besar terbagi menjadi 5 jenis yaitu plagioklas + kuarsa, *hornblende*, ortoklas, muskovit, dan *garnet*. Gambar 7. merupakan foto mineral batuan yang telah dilakukan deskripsi secara *quicklook* menggunakan foto pada batuan.



**Gambar 7.** Mineral batuan pada dataset: (a) *hornblende*, (b) plagioklas + kuarsa, (c) *garnet*, (d) ortoklas, (e) muskovit

Daerah penelitian mencakup databor WPD02, WPD04, 1s deskripsi *quick-look* mineral yang dilengkapi oleh informasi litologi geologi mengenai daerah penelitian.

**Tabel 2.** Informasi geologi pada klaster dengan jumlah 2

2 KLASTER				
Klaster	WPD02	WPD04	WPD07	Informasi
0	didominasi oleh plg + qtz (65%). Hb (25%), gnt (12%), ort (12%), mus (11%)	didominasi oleh gnt (45%). Plg + qtz (35%), hb (8%), ort (7%), mus (5%)	didominasi oleh plg + qtz (70%). ort (15%), gnt (10%), hb (10%), mus (5%)	Didominasi oleh <i>Felsic gneiss</i> dan <i>amphibolite</i> . Terdapat <i>intermediate gneiss</i> , <i>mafic gneiss</i> , <i>amphibolite</i> (WPD4), <i>mafic gneiss</i> , <i>amphibolite</i> , <i>felsic volcanic breccia</i> (WPD02), dan <i>amphibolite</i> , <i>felsic volcanic</i> , CLR (WPD7)
1	didominasi oleh plg + qtz (45%). Hb (35%), mus (8%), ort (6%), gnt (6%)	didominasi oleh plg + qtz (50%). Hb (35%), mus (10%), gnt (5%)	didominasi oleh plg + qtz (45%). Hb (35%), ort (10%), mus (5%), gnt (5%)	Didominasi oleh CLR, <i>amphibolite</i> . Terdapat <i>shear zone</i> , <i>mafic gneiss</i> (WPD04), <i>gabbro</i> , <i>felsic gneiss</i> (WPD02), <i>mafic gneiss</i> (WPD07)

**Tabel 3.** Informasi geologi pada klaster dengan jumlah 3

3 KLASTER				
Klaster	WPD02	WPD04	WPD07	Informasi
0	didominasi oleh plg + qtz (40%). Hb (25%), gnt (12%), ort (12%), mus(11%)	didominasi oleh plg + qtz(45%). Hb (25%), gnt (10%), ort (10%), mus(10%)	didominasi oleh plg + qtz (50%). Hb (25%), gnt (8%), ort (8%), mus (9%)	didominasi oleh CLR dan <i>amphibolite</i> . Terdapat <i>dolerite</i> (WPD4), <i>gabbro</i> (WPD02), dan <i>mafic gneiss</i> (WPD7)



**3 KLAS TER**

Klaster	WPD02	WPD04	WPD07	Informasi
1	didominasi oleh plg + qtz (60%). ort (40%)	didominasi oleh plg + qtz (60%). sedikit gnt (14%), ort (13%), mus (13%)	didominasi oleh plg + qtz (55%), mus (30%), hb (15%)	didominasi oleh felsic gneiss, sedikit terdapat CLR dan amphibolite (WPD4 dan 7)
2	didominasi oleh plg + qtz (60%). Hb (20%), ort (10%), mus(20%)	didominasi oleh plg + qtz(60%). Hb (14%), gnt (13%), ort (13%)	didominasi oleh plg + qtz (55%). Hb (35%), gnt(10%)	terdiri dari CLR, felsic gneiss, intermediate gneiss, amphibolite, shear zone

**Tabel 4.** Informasi geologi pada klaster dengan jumlah 4

**4 KLAS TER**

Klaster	WPD02	WPD04	WPD07	Informasi
0	didominasi oleh plg + qtz (50%). Hb (30%), gnt (5%), ort (10%), mus (5%)	didominasi oleh plg + qtz (60%). Hb (30%), gnt (5%), mus(5%)	didominasi oleh plg + qtz (60%). Hb (25%), ort (10%), gnt(5%)	didominasi oleh CLR dan amphibolite, terdapat juga dolerite (WPD4), gabbro (WPD2), shear zone (WPD7)
1	didominasi oleh hb (50%). Plg + qtz (35%), mus (10%), ort (5%)	didominasi oleh gnt (65%). Plg + qtz (15%), mus (10%), ort (10%)	didominasi oleh plg + qtz (45%). Hb (35%), mus (20%)	didominasi oleh felsic gneiss, terdapat intermediate gneiss (WPD4), mafic gneiss (WPD7), dan CLR (WPD2)
2	didominasi oleh plg + qtz (45%). Hb (25%), gnt (20%), ort(5%),mus(5%)	terdiri dari plg + qtz (55%), gnt (30%). Sedikit terdapat hb (15%)	didominasi oleh plg + qtz (55%). Hb (25%), gnt (10%), ort (5%), mus(5%)	didominasi oleh CLR dan amphibolite, terdapat sedikit felsic gneiss (WPD2 dan 4), felsic volcanic (WPD7)
3	didominasi oleh plg + qtz (80%). sedikit ort (7%), hb(8%)	didominasi oleh plg + qtz (85%). sedikit ort (5%), gnt(7%), hb(3%)	didominasi oleh plg + qtz (85%), ort (7%), hb (8%)	menurut litologi lokal, didominasi oleh felsic gneiss, kemudian terdapat felsic volcanic (WPD07), intermediate dan mafic gneiss (WPD04), CLR dan mafic gneiss (WPD02)

**Tabel 5.** Informasi geologi pada klaster dengan jumlah 5

**5 KLAS TER**

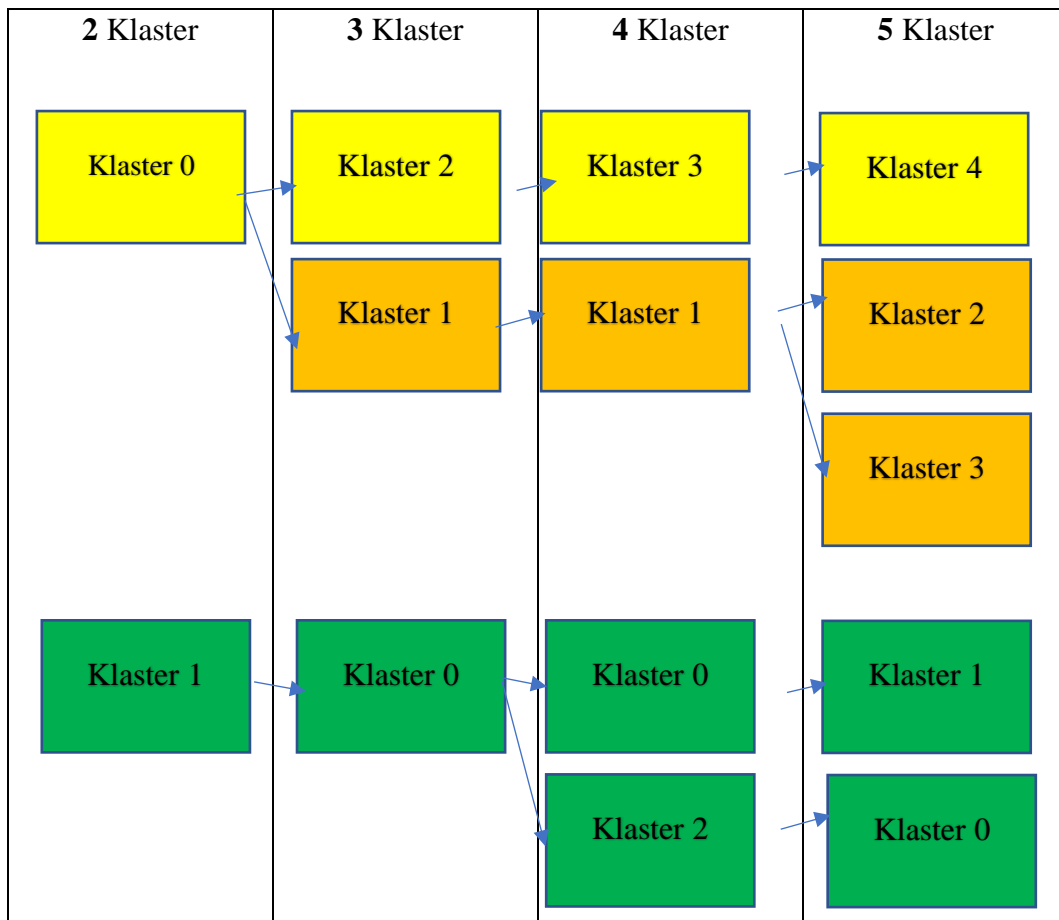
Klaster	WPD02	WPD04	WPD07	Informasi
0	didominasi oleh plg + qtz (40%). Hb (30%), gnt (10%), ort (10%), mus (10%)	didominasi oleh plg + qtz (40%). Hb (25%), gnt (25%), ort (10%).	didominasi oleh plg + qtz (50%). Hb (45%), ort (5%)	didominasi oleh CLR dan amphibolite, sedikit mengandung intermediate gneiss
1	didominasi oleh plg + qtz (55%). hb (25%), gnt (10%), ort (10%), mus (10%)	didominasi oleh plg + qtz (60%). hb (26%), gnt(17%), mus (17%).	didominasi oleh plg + qtz (50%), hb (40%), ort (10%)	didominasi oleh CLR dan amphibolite.
2	didominasi oleh plg + qtz (55%). Hb (30%), mus(15%)	didominasi oleh plg + qtz (35%). hb (25%), gnt (25%), ort (8%), mus (8%).	didominasi oleh plg + qtz (50%). Hb (25%), mus(15%), ort (10%)	didominasi oleh felsic gneiss, amphibolite dan CLR
3	didominasi oleh plg + qtz (55%), mus (45%)	didominasi oleh gnt + qtz (55%). plg (35%), ort (10%)	didominasi oleh plg + qtz (50%) dan mus(50%)	felsic gneiss, CLR
4	didominasi oleh plg + qtz (50%). Hb (40%), ort (10%)	didominasi oleh plg + qtz (40%). gnt (30%), ort (30%)	didominasi oleh plg + qtz (75%). Hb (20%), gnt(5%)	dominan felsic gneiss dan CLR, sedikit amphibolite



**3.4 Evaluasi Klaster**

Berdasarkan hasil klasterisasi yang telah dilakukan dengan menginput informasi geologi berupa mineral dan litologi batuan, menunjukkan bahwa klaster 1 pada 2 klaster, klaster 0 pada 3 klaster, klaster 0 dan 2 pada 4 klaster, dan klaster 0 dan 1 pada 5 klaster memiliki kesamaan ciri yaitu didominasi oleh mineral plagioklas + kuarsa dan *hornblende* dengan litologi batuan dominan CLR dan *amphibolite*. Kemudian klaster 0 pada 2 klaster, klaster 1 dan 2 pada 3 klaster, klaster 1 dan 3 pada 4 klaster, dan klaster 2, 3, dan 4 pada 5 klaster secara umum memiliki kemiripan ciri yaitu didominasi oleh mineral plagioklas + kuarsa dengan litologi yang didominasi oleh *felsic gneiss* dan amphibole serta sedikit CLR. Untuk menentukan jumlah klaster yang optimal, dibutuhkan pertimbangan mengenai perbedaan informasi geologi pada setiap klaster agar dalam 1 klaster, memiliki suatu ciri khusus yang berbeda dengan klaster lainnya secara geologi. 3 klaster merupakan jumlah klaster yang paling optimal karena memiliki perbedaan informasi geologi serta sehingga setiap klaster dapat merepresentasikan informasi geologi yang terkandung di dalamnya antara klaster 0, 1, dan 2.

**Tabel 6.** Informasi kesamaan ciri pada klaster



**4. KESIMPULAN**

Hasil klasterisasi yang dilakukan menggunakan metode PCA dan *K-means*, serta informasi geologi mengenai klaster, dapat disimpulkan bahwa klaster dengan jumlah k=3 merupakan jumlah klaster yang paling optimal dan ideal karena memiliki informasi geologi berupa data mineral dan litologi batuan yang dapat dibedakan cirinya antara klaster 0, klaster 1, dan klaster 2 serta memiliki batas antar klaster yang tegas. Variasi jumlah 2 klaster tidak dipilih sebagai klaster paling optimal karena setiap klaster mengandung informasi yang berlebihan sehingga kedua klaster tersebut bersifat umum dan tidak merepresentasikan informasi klaster. Sedangkan, variasi jumlah 4 klaster tidak dipilih juga karena klaster 0 dan klaster 2 memiliki informasi yang serupa sehingga masih dapat digabungkan menjadi satu. Variasi jumlah 5 klaster juga memiliki informasi klaster yang sama antara klaster 0 dan 1 klaster 2.

**REFERENCES**

[1] Klaus Schwab, *The Fourth Industrial Revolution*, World Econ. Geneva, 2016.  
 [2] K. Parakh, S. Thakur, B. Chudasama, S. Tirodkar, A. Porwal, and A. Bhattacharya, "Machine learning and spectral techniques for lithological classification," *Multispectral, Hyperspectral, Ultraspectral Remote Sens. Technol. Tech. Appl. VI*, vol. 9880, no. April, p. 98801Z, 2016, doi: 10.1117/12.2223638.



- [3] A. Caté, L. Perozzi, E. Gloaguen, and M. Blouin, "Machine learning as a tool for geologists," *Lead. Edge*, vol. 36, no. 3, pp. 215–219, 2017, doi: 10.1190/tle36030215.1.
- [4] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Dev.*, vol. 44, no. 1–2, pp. 207–219, 1959, doi: 10.1147/rd.441.0206.
- [5] Andreas C. Müller and Sarah Guido, *Introduction to with Python Learning Machine*. O'Reilly Media, Incorporated, 2016.
- [6] J. W. G. Putra, "Pengenalan Konsep Pembelajaran Mesin dan Deep Learning," pp. 1–235, 2019.
- [7] A. Ali and C. Sheng-Chang, "Characterization of well logs using K-mean cluster analysis," *J. Pet. Explor. Prod. Technol.*, vol. 10, no. 6, pp. 2245–2256, 2020, doi: 10.1007/s13202-020-00895-4.
- [8] B. Savage, "Wheatley project, Annual report E70/2258 & E70/2339: Hampton Hill Mining Wheatley Project," Australia, 2007.
- [9] D. W. Coulter, P. D. Harris, L. M. Wickert, and X. Zhou, "Advances in Spectral Geology and Remote Sensing: 2008 – 2017," *Proc. Explor. 17 Sixth Decenn. Int. Conf. Miner. Explor.*, pp. 23–50, 2017.
- [10] L. Shu, K. McIsaac, G. R. Osinski, and R. Francis, "Unsupervised feature learning for autonomous rock image classification," *Comput. Geosci.*, vol. 106, no. March 2016, pp. 10–17, 2017, doi: 10.1016/j.cageo.2017.05.010.
- [11] T. S. Madhulatha, "An Overview on Clustering Methods," *IOSR J. Eng.*, vol. 2, no. 4, pp. 719–725, 2012.
- [12] C. Klein, *The 22nd Edition of the Manual of Mineral Science*, 22nd ed. New York: John Wiley & Sons, 2002.
- [13] L. Hassan, *Metamorphosed Vms Mineralization At Wheatley, Southwest At Wheatley, Southwest*, no. August. 2017.