PENGELOMPOKAN DATA KELAS DESA BERDASARKAN DATA *LETTER - C* MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING*

Asrul Hidayahtul Kusna[1], Indyah Hartami Santi[2], Filda Febrinita[3]

[1,2,3] Teknik Informatika, Teknologi Informasi, Universitas, Universitas Islam Balitar
Jl Majapahit No. 04 A, Blitar, Jawa Timur, Indonesia
[1]asrulhidayah682@gmail.com, [2]indyahartamisanti@gmail.com, [3]febrinitafilda80@gmail.com

Abstrak

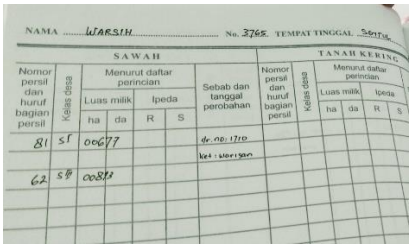
Data *Letter - C* adalah dokumen pertanahan yang digunakan sebagai acuan untuk membuktikan kepemilikan tanah. Dalam melakukan pengelompokan jenis tanah atau kelas desa di Desa Pandanarum masih dilakukan secara manual sehingga berakibat pada kurang efektifnya dalam pengelolaan sumber daya pertanahan. Untuk itu dilakukan pengelompokan kelas desa dengan menggunakan Algoritma *K – Means Clustering* dengan penentuan jumlah *cluster* menggunakan metode *Elbow*. Pengujian Algoritma dilakukan melalui metode *Davies Bouldin Index*. Proses penelitian dilakukan dengan menggunakan data *Letter – C* buku ketiga, dengan jumlah data sebanyak 429 data. Dari penelitian yang dilakukan. menghasilkan pengelompokan data kelas desa sebanyak 7 *cluster*. *Cluster* 1 memiliki sebaran kelas desa sebanyak 71 data yang terdapat pada blok 19,20,21,22,23 dan 24, sementara pada *cluster* terakhir, yaitu *cluster* ke 7, sebaran kelas desa sebanyak 84 data yang terdapat pada blok 12,13,14,15,16,17. Selanjutnya untuk proses pengujian dengan metode DBI, menghasilkan nilai 0,854, sehingga dapat disimpulkan pengelompokan data kelas desa dalam 7 cluster adalah kelompok tepat atau terbaik. Berdasarkan hasil penelitian tersebut dapat disimpulkan bahwa pengelompokan data kelas menggunakan algoritma *K-Means* di Desa Pandanarum dapat meningkatkan pengelolaan sumber daya pertanahan.

Kata kunci : CRISP-DM, *k-means*, *clustering*, *elbow*, DBI, *letter-c*.

1. Pendahuluan

Pertumbuhan data pada era digital saat ini menunjukkan perkembangan yang sangat pesat dan melibatkan bermacam - macam jenis informasi. Salah satunya adalah informasi terkait data pertanahan tingkat desa atau yang disebut *letter – C*. Pandanarum merupakan sebuah desa yang berada dibagian selatan Kabupaten Blitar tepatnya di Kecamatan Sutojayan. Desa Pandanarum terdiri dari 44 RT dan 10 RW yang berada di tiga (3) dusun yaitu Pandanarum, Klampok, dan Sentul dengan jumlah penduduk sebanyak 8.300 jiwa yang terbagi kedalam 2.561 KK dan memiliki luas wilayah sebesar 369 hektar(Santi et al., 2023).

Pengelolaan data pertanahan di Desa Pandanarum menjadi semakin penting untuk proses pembangunan lokal yang efisien dan berkelanjutan. Dokumen *letter – C* yang terdapat didesa merupakan data vital yang tidak dapat diperbaharui dan digantikan apabila terjadi kerusakan atau hilang(Setiawan et al., 2022). Buku *letter – C* dapat digunakan untuk membantu menetukan nilai jual dan beli suatu tanah, serta disebut juga sebagai pepel yang dipergunakan oleh petugas pemungut pajak untuk dijadikan sebagai patokan untuk pembayaran pajak di zaman pendudukan kolonial Belanda (Soepadi & Widodo, 2021). Patokan yang dijadikan acuan untuk menentukan besarnya nilai pajak yang harus dibayarkan dapat ditentukan dari jenis tanah yang tercantum dalam kelas desa yang terdapat dalam buku *letter – C*. Kelas desa merupakan tipe tanah berdasarkan keadaan lingkungan tanah yang dibagi menjadi dua (2) yaitu daratan dan sawah, kemudian dari kedua jenis tersebut masih dibagi menjadi tujuh (7) kategori kelas desa, form buku *letter – c* seperti pada gambar 1 berikut :



Gambar 1. Form Buku *Letter - C*

Survey observasi dan wawancara yang dilakukan menemukan permasalahan pada pengelompokan data kelas desa yang tepat pada blok tanah di Desa Pandanarum. Sehingga hal ini mengakibatkan kesulitan untuk melakukan pengelolaan sumber daya pertanahan, identifikasi wilayah yang memerlukan perhatian khusus serta perencanaan tataruang yang lebih baik.

Dari permasalahan yang timbul diatas diperlukan pengelompokan data sebaran kelas desa berdasarkan blok tanah di Desa Pandanarum. Pengelompokan data kelas desa berdasarkan buku *Letter–C* dilakukan menggunakan metode *Elbow* untuk menentukan jumlah *cluster*, menerapkan

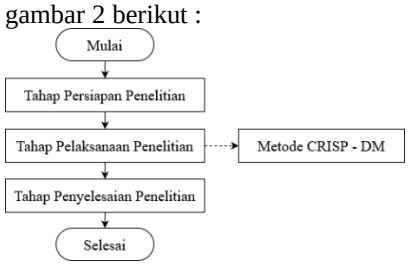
Algoritma *K–Means Clustering* untuk pengelompokan data, dan metode *Davies Bouldin Index* untuk pengujian dari perhitungan *K – Means*. Penerapan algoritma *k-means* dalam penelitian ini diharapkan dapat memberikan sebuah informasi yang lebih jelas mengenai pola distribusi kepemilikan tanah atau lahan, jenis penggunaan lahan, dan karakteristik pertanahan lainnya diberbagai kelas desa. Kelebihan dari algoritma *k – means* ini adalah sangat mudah dipahami dan diimplementasikan, dapat digunakan untuk data yang memiliki banyak variabel, dan pada perhitungan ulang *centroid* sebuah *instance* dapat mengubah *cluster* (Novia et al., 2020).

Dalam penelitian terdahulu yang dilakukan Santi et al., (2023) menghasilkan sebuah rancangan rekayasa sistem pengumpulan, pencarian, dan pemetaan data *letter – C* di Desa Pandanarum. Penelitian yang diselesaikan oleh Merly et al., (2024) menghasilkan pengelompokan data status pertanahan menggunakan algoritma *k-medoids*, kedalam 3 *cluster*. Penelitian yang dilakukan oleh Koto et al., (2020) menghasilkan cara pengelompokan pemegang sertifikat hak milik atas tanah, mengelompokkan pemegang sertifikat hak atas tanah yang pantas untuk diberikan target penerbitan sertifikat masal dan tanpa dipungut biaya dengan menggunakan algoritma *k – means*. Kemudian penelitian yang dilakukan oleh Supriyadi et al., (2021) menunjukkan bahwa berdasarkan nilai keabsahan algoritma *k – means* lebih relevan untuk diterapkan dalam pembuatan sistem aplikasi pengelompokan armada kendaraan berbasis *website* dengan nilai DBI lebih rendah dari *k–medoids*. Dengan evaluasi yang dilakukan menghasilkan persentasi kesesuaian sebesar 97% dengan menggunakan aplikasi *Rapidminer* dan perhitungan manual.

Berdasarkan identifikasi masalah serta beberapa kajian penelitian yang telah dilaksanakan, maka dilakukan penelitian terkait "Pengelompokan Data Kelas Desa Berdasarkan Data *Letter - C* Menggunakan Algoritma *K-Means Clustering*". Penelitian ini ditujukan untuk memberikan kontribusi dalam pemetaan dan perencanaan pembangunan berkelanjutan serta meningkatkan efektivitas pengelolaan pertanahan desa secara keseluruhan.

2. Metode

Pada penelitian ini terdapat beberapa tahapan atau langkah dalam proses penelitian, Adapun tahapan tersebut dapat dilihat pada *flowchart* seperti pada gambar 2 berikut :



Gambar 2. *Flowchart* Tahapan Penelitian

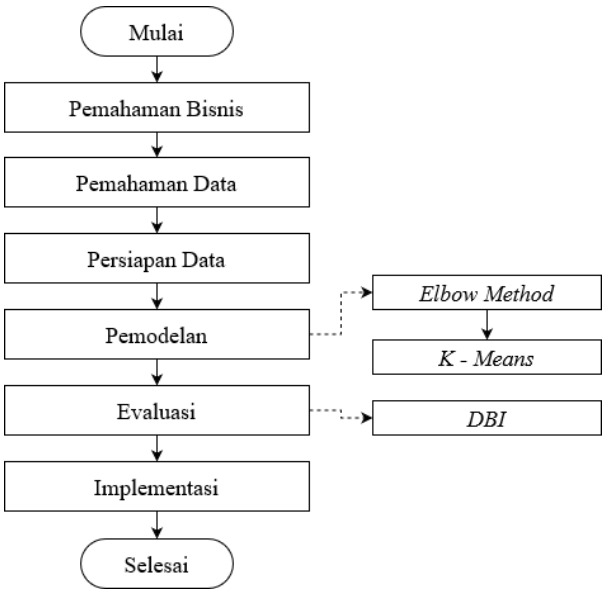
Berdasarkan flowchart pada gambar 2 terdapat 3 tahapan dalam penelitian ini, tahap yang pertama adalah tahap Persiapan Penelitian, kemudian Tahap Pelaksanaan Penelitian yang didalamnya terdapat Metode CRISP – DM, dan yang terakhir adalah tahap Penyelesaian Penelitian

2.1 Tahap Persiapan Penelitian

Tahapan persiapan dalam penelitian ini dilakukan dengan melakukan observasi ke Desa Pandanarum. Melakukan koordinasi dan konsultasi dengan Kepala Desa Pandanarum untuk mengetahui keadaan dan kondisi dokumen atau objek yang akan diteliti. Serta melakukan pengkajian teori – teori terkait data *Letter* – C, penentuan jumlah cluster menggunakan metode *Elbow*, pengelompokan data dengan algoritma *K – Means*, dan pengujian algoritma menggunakan metode DBI.

2.2 Tahap Pelaksanaan Penelitian

Tahapan pelaksanaan penelitian ini menerapkan Metode CRISP – DM atau *Cross Industry Standart Process for Data Mining*. Metode ini merupakan metodologi yang digunakan dalam pengolahan data mining serta analisis data (Sulianta, 2023). Metode ini digunakan untuk melakukan analisis data serta membuat sebuah model prediktif, dan metode ini juga dianggap sebagai prinsip panduan paling relevan dan komprehensif yang berfungsi untuk melaksanakan proyek analitik (Nasution & Fatonah, 2023). Proses data mining CRISP - DM berbentuk lingkaran yang memiliki beberapa tahap dan semua berputar Kembali. Perputaran berfungsi untuk memberikan pemahaman baru yang terdapat pada tahap sebelumnya (Sumarauw, 2022). Metode ini memiliki 6 tahapan, seperti pada gambar 3 berikut :



Gambar 3. *Flowchart* Tahapan CRISP – DM

2.2.1 *Business Understanding*

Tahap ini merupakan proses untuk melakukan pemahaman tentang masalah terkait dengan pengolahan data.

2.2.2 *Data Understanding*

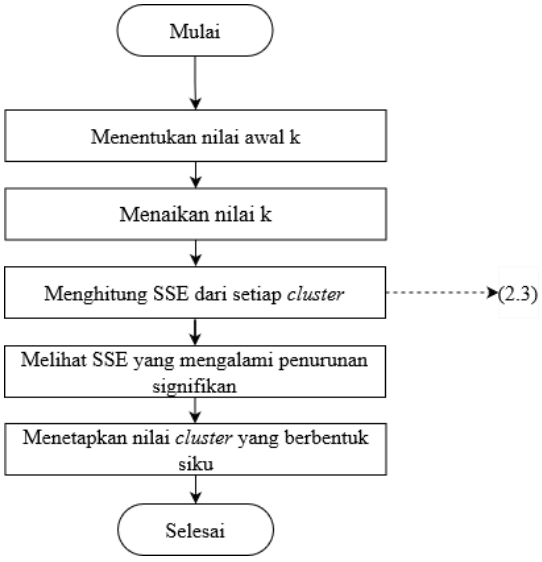
Pada tahap kedua diperlukan data informasi untuk melakukan penelitian, dimana data tersebut tercantum dalam buku Register C dan juga peta tanah desa pandanarum.

2.2.3 *Data Preparation*

Pada tahap ini dilakukan proses pembersihan data atau *cleaning* data, untuk menghilangkan data – data yang tidak lengkap dan tidak digunakan, sehingga data yang ditampilkan adalah data yang sesuai dengan kebutuhan dan digunakan dalam proses pemodelan data (*modelling*).

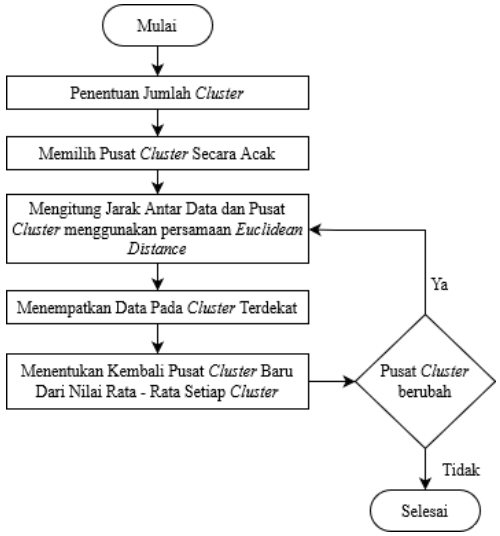
2.2.4 *Modelling*

Pada tahap pemodelan data dilakukan penerapan metode *elbow* yang merupakan Teknik optimalisasi untuk menentukan jumlah *cluster*. Dengan melakukan perbandingan persentase antar *cluster* sehingga membentuk sudut siku terhadap suatu titik. Dari hasil persentase yang tidak sama pada setiap *cluster* dapat ditunjukkan dengan sebuah *grafik* yang dijadikan sebagain sumber informasi (Jollyta et al., 2021). Menurut Yuliana Sari et al., (2022) langkah – langkah metode *elbow* seperti pada gambar 4 berikut :



Gambar 4. *Flowchart* Metode *Elbow*

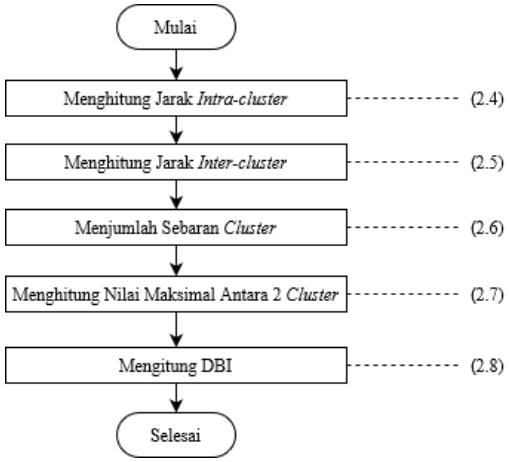
Setelah mendapatkan jumlah *cluster* yang tepat selanjutnya dilakukan perhitungan *data mining*. Pada penelitian ini pengelompokan dilakukan dengan menerapkan algoritma *k-means clustering*. Algoritma *K – Means* adalah operasi pengelompokkan pada set data tanpa label serta merupakan algoritma pembelajaran tanpa pengawasan (DIng et al., 2021). Menurut Febrinita et al., (2023) Algoritma ini memiliki tujuh (7) tahap untuk melakukan pengelompokan data, seperti pada gambar 5 berikut :



Gambar 5. *Flowchart* Tahapan K-Means *Clustering*

2.2.5 *Evaluation*

Tahap berikutnya adalah tahap evaluasi, dengan menerapkan metode *Davies Bouldin Index* atau DBI untuk memaksimalkan jarak antar sampel dalam satu kelompok atau kelas, sehingga DBI digunakan untuk menguji validasi data dalam setiap kelompok (Mustika et al., 2021). Metode ini bekerja dengan mempertimbangkan kesamaan rata – rata antara setiap *cluster* dan mencari *cluster* yang serupa (Fhadli & Tempola, 2020). DBI memiliki beberapa tahapan dalam melakukan evaluasi, seperti pada gambar 6 berikut :



Gambar 6. *Flowchart* Tahapan DBI

2.2.6 *Deployment*

Pada tahap ini merupakan tahap Deployment atau Implemtasi, setelah dilakukan perhitungan manual dan evaluasi selanjutnya dilakukan implementasi algoritma *K-Means* kedalam bahasa

pemrograman *phyton* dengan menampilkan hasil *console*.

2.3 Tahap Penyelesaian

Pada penelitian ini pengelompokan kelas desa dilakukan dengan mengguanakan metode *elbow* untuk menentukan *cluster* terbaik. Kemudian dari *cluster* yang didapatkan dilakukan penerapan algoritma *K – Means* untuk melakukan pengelompokan kelas desa dalam sebaran blok berdasarkan data *letter – C*. Selanjutnya dilakukan evaluasi atau pengujian hasil *K – Means* dengan menerapkan metode *Davies Bouldin Index* (DBI).

3. Hasil dan Pembahasan

3.1 Hasil

Hasil dari penelitian yang dilakukan adalah pengelompokan data kelas desa dengan menerapkan Algoritma *K – Means Clustering*. Pada pengelompokan ini data yang digunakan sebanyak 429 data yang diperoleh dari buku register – C atau *Letter – C* versi ketiga Desa Pandanarum, seperti pada tabel 1 berikut :

Tabel 1. Data *Letter - C* buku ketiga

No	Nomor Register C	Nama Pemilik	Nomor Blok	Nomor Persil	Luas Tanah
1	3609	Samirin/ Lisawati			
2	3610	Samilah			
...	...	...	...	...	...
428	3959	Sumaji - Winarsih	17	76	798
429	3960	Nur Paidah - Sutrisno	17	76	765

Berdasarkan tabel data *letter-c* buku ketiga dilanjutkan proses *cleaning* data untuk menghapus data – data yang tidak sesuai dengan atribut yang digunakan dalam perhitungan algoritma *K – Means*, hasil proses seperti pada tabel 2 berikut :

Tabel 2. Data hasil *cleaning data*

No	Nomor Register C	Nama Pemilik	Nomor Blok	Nomor Persil	Luas Tanah
1	3611	Moch Solikin - Manidah	23	121	284
2	3612	Rolis Pratikno	12	47	923
...	...	...	...	...	...
406	3959	Sumaji - Winarsih	17	76	798
407	3960	Nur Paidah - Sutrisno	17	76	765

Berdasarkan hasil *cleaning data* pada tabel 2, selanjutnya data akan digunakan untuk melakukan proses perhitungan, berikut adalah penjabaran dari perhitungan yang dilakukan dalam penelitian ini.

3.1.1 Perhitungan Metode *Elbow*

a. Menentukan Nilai Awal K

Pada tahap awal penentuan nilai awal *cluster* k adalah dua (2) sebagai nilai *cluster* terendah dalam simulasi perhitungan pada penelitian yang dilakukan.

b. Menaikkan Nilai K

Langkah kedua adalah menaikkan nilai *cluster* sampai dengan ditentukan jumlah *cluster* terbaik

c. Menghitung Nilai SSE hingga *cluster* yang ditentukan

Langkah selanjutnya adalah menghitung nilai SSE, pada proses ini data yang digunakan merupakan data hasil dari *cleaning* data, kemudian dilakukan normalisasi data. Normalisasi data merupakan proses pengaturan nilai dalam suatu dataset menjadi rentang yang kebih terstandarisasi atau normal. Normalisasi dilakukan sebelum proses perhitungan jarak sebuah data (Arhami & Nasir, 2020). Norrmalisasi data dilakukan pada setiap atribut yang digunakan yaitu Nomor Blok, Nomor Persil, dan Luas Milik. Proses normalisasi dilakukan dimana nilai – nilai dalam data dibagi dengan nilai data maksimum dan nilai data minimum dari data tersebut. Menurut Mulaab, (2021) normalisasi data dilakukan mengikuti persamaan 1 berikut :

$$f(x) = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad \text{.....(1)}$$

Tabel 3. Data Maksimum dan Minimum dataset

	Nomor Blok	Nomor Persil	Luas Milik
Nilai Minimun	1	1	42
Nilai Maksimum	29	158	3097

Setelah ditentukan nilai maksimum dan minimum dataset pada tabel 3, selanjutnya dilakukan normalisasi data pada setiap atribut yang digunakan seperti pada tabel 4 berikut :

Tabel 4. Data Normalisasi

No	Nomor Register C	Nama Pemilik	Nomor Blok	Nomor Persil	Luas Tanah
1	3611	Moch Solikin - Manidah	0,786	0,764	0,079
2	3612	Rolis Pratikno	0,393	0,293	0,288
...	...	...	...	...	...
406	3959	Sumaji - Winarsih	0,571	0,478	0,247
407	3960	Nur Paidah - Sutrisno	0,571	0,478	0,237

Berdasarkan data normalisasi pada tabel 4, dilanjutkan proses perhitungan Nilai SSE dengan mengikuti persamaan 2 berikut :

$$SEE = \sum_{K=1}^{K} \sum_{x_i} \left\| x_{ij-} c_k \right\|_2^2 \text{ ....(2)}$$

Dalam penelitian ini perhitungan dilakukan dengan nilai awal k = 2, dan dilakukan penaikan nilai k sampai dengan 9, menggunakan data *centroid* awal yang diambil secara acak, data *centroid* awal seperti pada tabel 5 berikut :

Tabel 5. Data *centroid* awal Perhitungan SSE 2 *cluster*

	Nomor Blok	Nomor Persil	Luas Milik	
Centroid 1	0,786	0,764	0,079	data ke 1
Centroid 2	0,786	0,783	0,182	data ke 81

Berdasarkan data *centroid* pada tabel 5, dilakukan perhitungan nilai SSE 2 cluster seperti pada tabel 6 berikut :

Tabel 6. Perhitungan SSE 2 *Cluster*

No	Nomor Register	*Cluster*		Jarak Terdekat	Inersia
		C1	C2		
1	3611	0,000	0,105	0,000	0,000
2	3612	0,648	0,637	0,637	0,406
...	...	...	...	...	...
406	3959	0,395	0,379	0,379	0,144
407	3960	0,391	0,377	0,377	0,142
Nilai SSE					143,440

Perhitungan nilai SSE terus dilakukan sampai dengan nilai k = 9 atau C9, dan mengambil data centroid secara acak.

d. Hasil Nilai SSE *Cluster* Yang Mengalami Penurunan Secara Signifikan

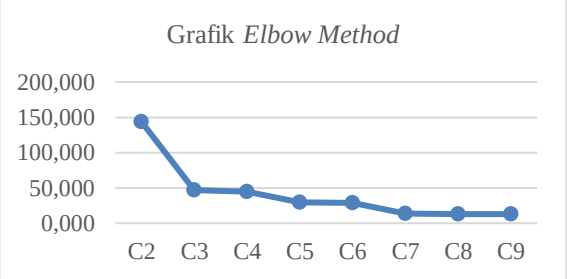
Setelah melakukan proses perhitungan SSE hingga jumlah *cluster* yang ditentukan, selanjutnya melakukan perhitungan selisih pada setiap nilai SSE untuk mengetahui selisih penurunan nilai yang digunakan sebagai acuan penentuan *cluster*. Seperti hasil yang terdapat pada tabel 7 berikut :

Tabel 7. Hasil Perhitungan Nilai SSE

C/k	Nilai SSE	Jarak	Keterangan
2	143,440	143,440	C2
3	47,125	96,314	Selisih C2 ke C3
4	44,723	2,402	Selisih C3 ke C4
5	29,215	15,508	Selisih C4 ke C5
6	28,882	0,333	Selisih C5 ke C6
7	13,309	15,573	Selisih C6 ke C7
8	13,126	0,184	Selisih C7 ke C8
9	12,869	0,256	Selisih C8 ke C9

Berdasarkan hasil perhitungan nilai SSE pada tabel 7 diatas terdapat dua nilai SSE yang mengalami penurunan signifikan, yaitu pada *cluster* ke 3 dan ke 7. Pada *cluster* ke 3 nilai SSE selanjutnya masih belum stabil dan masih mengalami penurunan angka, dan pada *cluster* ke 7 nilai SSE selanjutnya mengalami penurunan yang relatif stabil.

e. Menetapkan Nilai Untuk *Cluster* Dari Grafik Yang Berbentuk Siku



Gambar 7. Grafik Hasil Perhitungan Nilai SSE

Dari hasil grafik nilai cluster pada gambar 7, nilai yang diambil untuk dijadikan sebagai *cluster* optimal dalam metode *elbow* terletak pada *cluster* ke 7, dengan titik membentuk siku dan terdapat penurunan yang cukup signifikan pada 2 titik selanjutnya.

Sehingga pada penelitian ini pengelompokan dilakukan menjadi 7 cluster, dengan C1 adalah tanah daratan dekat dengan jalan desa, C2 tanah daratan agak jauh dari jalan desa, C3 tanah daratan jauh dari jalan desa, C4 tanah daratan sangat jauh dari jalan desa, C5 tanah sawah dekat dengan pemukiman, C6 tanah sawah jauh dengan pemukiman, C7 tanah sawah sangat jauh dari pemukiman.

3.1.2 Perhitungan *K – Means Clustering*

a. Menentukan Jumlah *Cluster*

Pada penelitian ini jumlah *cluster* ditentukan menggunakan metode *elbow*. Dari hasil perhitungan *elbow* sebelumnya jumlah *cluster* yang optimal adalah 7. Sehingga pada perhitungan *K – Means* untuk pengelompokan data kelas desa di Desa Pandanarum menggunakan 7 *cluster*.

b. Memilih Pusat *Cluster* Secara *Random*

Pusat *cluster* atau *centroid* awal yang digunakan untuk perhitungan *K-Means* ditentukan secara acak, yaitu data ke 1, ke 90, ke 182, ke 280, ke 356, ke 390, dan data ke 407.

c. Menghitung Jarak Dengan Persamaan *Euclidean Distance*

Euclidean Distance merupakan persamaan perhitungan jarak yang sering digunakan, karena mendukung perhitungan *cluster* dan mudah dipahami. Persamaan ini menghasilkan perhitungan jarak terpendek diantara 2 titik yang dilakukan hingga mendapatkan nilai pengelompokan yang konvergen (Jollyta et al., 2021). Menurut Febrinita et al., (2023) perhitungan *Euclidean Distance* mengikuti persamaan 3 berikut :

$$d(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + \ldots + (x_{ki} - x_{kj})^2} \ldots\ldots\ldots(3)$$

Dalam tahap ini dilakukan perhitungan jarak data setelah dilakukan normalisasi ke *centroid* yang ditentukan secara acak pada Iterasi ke – I. Tabel perhitungan Iterasi I seperti pada tabel 8 berikut :

Tabel 8. Perhitungan Jarak Data Iterasi 1

No	Nomor Register	d(C1)	…	d(C7)	Jarak Terdekat	Hasil
1	3611	0,000	…	0,391	0,000	C1
2	3612	0,648	…	0,262	0,194	C2
…	…	…	…	…	…	…
406	3959	0,395	…	0,011	0,011	C7
407	3960	0,391	…	0,000	0,000	C7

d. Menentukan Kembali Pusat *Cluster* Baru dan Melakukan Perhitungan Jarak *Euclidean Distance*

Pada proses ini dilakukan penentuan pusat *cluster* baru dari nilai rata – rata setiap *cluster* dan melakukan penghitungan kembali jarak antar data dengan pusat *cluster* baru. Perhitungan terus dilakukan berulang hingga tidak terdapat perubahan pada titik pusat *cluster*. Data *centroid* baru yang diperoleh dari nilai rata – rata Iterasi I seperti pada tabel 9 berikut :

Tabel 9. Data *Centroid* dari Iterasi 1

	Nomor Blok	Nomor Persil	Luas Milik	
Centroid 1	0,744	0,708	0,093	*Cluster* 1 Iterasi 1
Centroid 2	0,357	0,207	0,288	*Cluster* 2 Iterasi 1
Centroid 3	0,649	0,586	0,221	*Cluster* 3 Iterasi 1
Centroid 4	0,139	0,096	0,136	*Cluster* 4 Iterasi 1
Centroid 5	0,929	0,927	0,204	*Cluster* 5 Iterasi 1
Centroid 6	0,627	0,546	0,520	*Cluster* 6 Iterasi 1
Centroid 7	0,500	0,360	0,121	*Cluster* 7 Iterasi 1

Berdasarkan data *centroid* dari iterasi 1 pada tabel 9, dilakukan perhitungan jarak data Iterasi ke – 2, seperti pada tabel 10 berikut :

Tabel 10. Perhitungan Jarak Data Iterasi 2

No	Nomor Register	d(C1)	…	d(C7)	Jarak Terdekat	Hasil
1	3611	0,071	…	0,497	0,071	C1
2	3612	0,578	…	0,210	0,093	C2
…	…	…	…	…	…	…
406	3959	0,327	…	0,187	0,136	C3
407	3960	0,322	…	0,180	0,134	C3

Perhitungan jarak data terus – menerus dilakukan sampai dengan tidak ditemukan perubahan pada jarak data antara 2 iterasi.

e. Hasil Dari Perhitungan *K -Means Clustering*

Dari perhitungan *K – Means* diatas didapatkan hasil pengelompokan data kelas desa sebanyak 7 *cluster*. Dengan hasil *cluster* 1 daratan dekat jalan desa terdapat sebaran kelas desa sebanyak 71 data yang terdapat pada blok 19, 20,21,22,23, dan 24. Sedangkan pada *cluster* terakhir (7) sawah jauh dari pemukiman terdapat sebaran kelas desa sebanyak 84 data yang terdapat pada blok 12,13,14,15,16, dan 17.

3.1.3 Perhitungan *Davies Bouldin Index*

Evaluasi Algoritma *K -Means Clustering* pada penelitian ini dilakukan dengan menerapkan metode *Davies Bouldin Index* (DBI). Data tersebut diambil dari hasil perhitungan Algoritma *K – Means Clustering*, seperti pada tabel 11 berikut :

Tabel 11. Dataset Perhitungan DBI

No	Nomor Registrasi	Nama Pemilik	Nomor Blok	Nomor Persil	Luas Milik	Hasil Cluster
1	3611	Moch Solikin - Manidah	0,786	0,764	0,079	C1
2	3612	Rolis Pratikno	0,393	0,293	0,288	C7
…	…	…	…	…	…	…
406	3959	Sumaji - Winarsih	0,571	0,478	0,247	C3
407	3960	Nur Paidah - Sutrisno	0,571	0,478	0,237	C3

Dataset berikutnya yang diperlukan untuk melakukan perhitungan DBI adalah pusat *cluster* atau *centroid* yang di gunakan untuk melakukan

perhitungan *K-Means* pada iterasi terakhir, seperti pada tabel 12 berikut :

Tabel 12. Data *Centroid* Iterasi Terakhir

Pusat	Nomor Blok	Nomor Persil	Luas Milik
Centroid 1	0,727	0,687	0,108
Centroid 2	0,292	0,168	0,130
Centroid 3	0,579	0,472	0,191
Centroid 4	0,053	0,061	0,158
Centroid 5	0,926	0,924	0,204
Centroid 6	0,614	0,528	0,490
Centroid 7	0,470	0,311	0,099

Berdasarkan dataset perhitungan DBI pada tabel 11 dan data *centroid* iterasi terakhir pada tabel 12, dilakukan perhitungan nilai DBI dengan langkah – langkah berikut :

a. Menghitung Jarak Intra – Cluster

Menghitung jarak *intra – cluster* untuk standar deviasi antar titik data dengan mengikuti persamaan 4 berikut :

$$s_i = \sqrt{\frac{1}{T_i}\sum_{j=1}^{T_i}|Xj - Ai|}\,.\,|Xj - Ai| \text{ ....(4)}$$

Berdasarkan pada tabel 11 dan tabel 12 dilakukan perhitungan jarak atara pusat *centroid* 1 dengan data *cluster* ke-1 sebanyak 71 data, pusat *cluster* 2 dengan data *cluster* ke-2 sebanyak 64 data, sampai dengan pusat *cluster* 7 dengan data *cluster* ke-7 sebanyak 84 data menggunakan persamaan 4. Kemudian dilakukan perhitungan jarak *intra-cluster* pada setiap *cluster*, dengan hasil seperti pada tabel 13 berikut :

Tabel 13. Hasil Perhitungan Jarak *Intra - Cluster*

	Jarak *intra – cluster*
S1	0,101
S2	0,100
S3	0,099
S4	0,143
S5	0,114
S6	0,173
S7	0,100

b. Menghitung Jarak *Inter – Cluster* Untuk Setiap Jarak Pasangan *Cluster*

Menghitung jarak *Inter – Cluster* dengan menggunakan titik pusat *cluster* atau *centroid* pada iterasi terkahir dalam perhitungan *K – Means*, dengan mengikuti persamaan 5 berikut :

$$M_{ij} = \sqrt{\|a_i - a_j\| \,.\, \|a_i - a_j\|} \text{ .....(5)}$$

Berdasarkan persamaan 5 dilakukan perhitungan jarak setiap titik pusat *cluster* menggunakan data pada tabel 12, perhitungan jarak *centroid* 1 dengan *centroid* 2, *centroid* 1 dengan *centroid* 3, *centroid* 1 dengan *centroid* 4, *centroid* 1 dengan *centroid* 5, *centroid* 1 dengan *centroid* 6, *centroid* 1 dengan *centroid* 7, sampai dengan *centroid* 6 dengan *centroid* 7, hasil perhitungan seperti pada tabel 14 berikut :

Tabel 14. Hasil Perhitungan *Intra – Cluster*

	M1	M2	M3	M4	M5	M6	M7
M1	0,000	0,678	0,274	0,921	0,323	0,429	0,456
M2	0,678	0,000	0,423	0,263	0,989	0,602	0,231
M3	0,274	0,423	0,000	0,668	0,569	0,305	0,216
M4	0,921	0,263	0,668	0,000	1,228	0,801	0,490
M5	0,323	0,989	0,569	1,228	0,000	0,580	0,771
M6	0,429	0,602	0,305	0,801	0,580	0,000	0,469
M7	0,456	0,231	0,216	0,490	0,771	0,469	0,000

c. Menjumlahkan sebaran *cluster* dan dibagi dengan jarak antara 2 *cluster*.

Langkah selanjutnya adalah menjumlahkan jarak setiap *cluster* dan membagi dengan jarak setiap titik pusat *cluster* dengan mengikuti persamaan 6 berikut :

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \text{ .....(6)}$$

Berdasarkan persamaan 6 dilakukan perhitungan rasio sebaran *cluster* menggunakan data pada tabel 13 dan tabel 14, dengan hasil seperti pada tabel 15 berikut :

Tabel 15. Hasil Perhitungan Sebaran 7 *Cluster*

	R1	R2	R3	R4	R5	R6	R7
R1	0,000	0,297	0,730	0,265	0,665	0,638	0,440
R2	0,297	0,000	0,472	0,925	0,217	0,454	0,867
R3	0,730	0,472	0,000	0,362	0,374	0,891	0,923
R4	0,265	0,925	0,362	0,000	0,209	0,394	0,496
R5	0,665	0,217	0,374	0,209	0,000	0,495	0,277
R6	0,638	0,454	0,891	0,394	0,495	0,000	0,581
R7	0,440	0,867	0,923	0,496	0,277	0,581	0,000

d. Mengukur atau menghitung nilai maksimal atara 2 *cluster*.

Tahap selanjutnya adalah menentukan nilai maksimal dari sebaran setiap *cluster* dengan mengikuti persamaan 7 berikut :

$R_i = \max R_{ij}$ .....(7)

Berdasarkan persamaan 7 dilakukan perhitungan nilai maksial menggunakan data pada tabel 15, dengan hasil seperti pada tabel 16 berikut :

Tabel 16. Hasil Perhitungan Nilai Maksimal

	R1	R2	R3	R4	R5	R6	R7
R1	0,000	0,297	0,730	0,265	0,665	0,638	0,440
R2	0,297	0,000	0,472	0,925	0,217	0,454	0,867
R3	0,730	0,472	0,000	0,362	0,374	0,891	0,923
R4	0,265	0,925	0,362	0,000	0,209	0,394	0,496
R5	0,665	0,217	0,374	0,209	0,000	0,495	0,277
R6	0,638	0,454	0,891	0,394	0,495	0,000	0,581
R7	0,440	0,867	0,923	0,496	0,277	0,581	0,000
R. MAX	**0,730**	**0,925**	**0,923**	**0,925**	**0,665**	**0,891**	**0,923**
R. TOTAL	5,981						

e. Menghitung Nilai DBI

Menghitung nilai DBI sebagai rata – rata ukuran kesamaan maksimal *cluster*, dengan mengikuti persamaan 8 berikut :

$DBI = \frac{1}{K} \sum_{i=i}^{K} R_i$ .....(8)

Nilai DBI dihitung dengan menggunakan data R Total pada tabel 16 yaitu 5,981 kemudian dibagi dengan jumlah *cluster* dalam penelitian ini yaitu 7, sehingga menghasilkan nilai DBI sebesar 0,854.

3.2 Pembahasan

3.2.1 Penentuan Jumlah *Cluster* Metode *Elbow*

Pada proses penentuan jumlah *cluster* dilakukan dengan metode *elbow*, menggunakan data sebanyak 407 data yang sebelumnya telah dilakukan *cleaning* data yang tidak lengkap dari data awal sebanyak 429. Kemudian dilakukan normalisasi data pada atribut nomor blok, nomor persil, dan luas milik, yang selanjutnya digunakan dalam perhitungan *elbow*. Langkah berikutnya dilakukan proses perhitungan *elbow*, langkah awal menghitung nilai SSE, dengan nilai awal k=2 dan dilakukan perhitungan sampai nilai k=9, data *centroid* yang digunakan diambil secara acak. Setelah didapatkan nilai SSE selanjutnya dilakukan perhitungan selisih pada setiap nilai untuk mengetahui selisih penurunan nilai yang digunakan sebagai acuan penentuan *cluster*. Langkah terakhir dalam penentuan jumlah *cluster* terbaik ditentukan dari grafik, yang menunjukan *cluster* terbaik terdapat pada titi ke 7, sehingga ditetapkan *cluster* terbaik adalah 7 *cluster* .

3.2.2 Perhitungan Algoritma *K–Means Clustering*

Pengelompokan data kelas desa dilakukan menggunakan algoritma *K – Means Clutering* dengan langkah awal menentukan jumlah *cluster* yang telah dihitung dengan menggunakan metode *elbow* yaitu 7 *cluster*. Kemudian memilih titik pusat *cluster* atau *centroid* awal secara acak, dalam penelitian ini *centroid* diambil dari data ke – 1, data ke – 90, data ke – 182, data ke – 280, data ke – 356, data ke – 390, dan data ke – 407. Setelah ditentukan nilai *centroid* awal selanjutnya dilakukan perhitungan jarak menggunakan persamaan *Euclidean Distance* untuk perhitungan Iterasi 1. Kemudian dari hasil *cluster* Iterasi ke – 1 digunakan untuk menentukan titik pusat *centroid* kedua, dengan mengambil nilai rata – rata pada setiap *cluster* untuk melakukan perhitungan jarak kembali pada Iterasi ke – 2. Perhitungan jarak dilakukan sampai dengan hasil *cluster* menghasilkan nilai yang sama dengan hasil *cluster* sebelumnya. Dalam penelitian ini perhitungan jarak dilakukan sampai dengan Iterasi ke – 13, pada Iterasi ke – 13 hasil *cluster* sama dengan Iterasi ke – 12, sehingga perhitungan dihentikan. Berdasarkan perhitungan Jarak yang dilakukan, menghasilkan sebaran *cluster* 1 tanah daratan dekat jalan desa sebanyak 71 data yang sebarannya terdapat pada blok 19,20,21,22,23 dan 24. Sebaran *cluster* 2 tanah daratan agak jauh jalan desa sebanyak 64 data yang sebarannya terdapat pada blok 7,8,9,10, dan 11. Sebaran *cluster* 3 tanah daratan jauh jalan desa sebanyak 75 data yang sebarannya terdapat pada blok 15,16,17,18, dan 19. Sebaran *cluster* 4 tanah daratan sangat jaut jalan desa sebanyak 55 data yang sebarannya terdapat pada blok 1,2,3,4, dan 5. Seberan *cluster* 5 tanah sawah dekat pemukiman sebanyak 36 data yang sebarannya terdapat pada blok 24,25,26,27,28 dan 29. Sebaran *cluster* 6 tanah sawah jauh pemukiman sebanyak 22 data yang sebarannya terdapat pada blok 13,14,16,17,18, 19,20,21,22 dan 26. Dan sebaran *cluster* 7 tanah sawah sangat jauh pemukiman sebanyak 84 data yang sebarannya terdapat pada blok 12,13,14,15,16 dan 17.

3.2.3 Evaluasi Perhitungan *K – Means* Menggunakan DBI (*Davies Bouldin Index*)

Evaluasi perhitungan *K – Means* dilakukan pengujian dengan menerapkan perhitungan DBI (*Davies Bouldin Index*). Dataset yang digunakan adalah data hasil *cluster* iterasi ke 13, dan dan data titik pusat *centroid* iterasi terakhir. Langkah pertama yang dilakukan dalam perhitungan DBI adalah menghitung jarak *intra-cluster* untuk digunakan sebagai standar deviasi antar titik data menggunakan persamaan 4. Kemudian langkah yang kedua adalah

menghitung jarak *inter-cluster* menggunakan data titik pusat *centroid* iterasi terkhir dalam perhitungan *K – Means* dengan persamaan 5. Langkah yang ketiga adalah menjumlahkan sebaran 7 *cluster* dan dibagi dengan jarak antara 2 *cluster*, dilakukan menggunakan persamaan 6. Langkah yang ke-empat adalah menentukan nilai maksimal dari setiap sebaran *cluster* yang dilakukan menggunakan persamaan 7. Langkah yang terakhir adalah menghitung DBI menggunakan persamaan 8, dengan hasil 0,854. Sehingga dapat disimpulkan hasil perhitungan *K–Means* adalah hasil yang tepat atau baik.

4. Kesimpulan dan Saran

Penerapan metode *elbow* dalam penentuan jumlah *cluster* menghasilkan 7 *cluster*. Penerapan algoritma *k-means clustering* menghasilkan sebaran *cluster* 1 sebanyak 71 data, *cluster* 2 sebanyak 64 data, *cluster* 3 sebanyak 75 data, *cluster* 4 sebanyak 55 data, *cluster* 5 sebanyak 36 data, *cluster* 6 sebanyak 22 data, *cluster* 7 sebanyak 84 data. Dan hasil dari evaluasi algoritma dengan metode DBI menghasilkan nilai 0,854 sehingga *cluster* yang dihasilkan tepat atau baik. Dari hasil penelitian ini dapat dilakukan penelitian lebih lanjut terkait pemilihan lebih spesifik terhadap atribut – atribut lain yang belum digunakan dalam penelitian ini. Sehingga dapat memaksimalkan hasil pengelompokan data kelas desa pada Registrasi C.

Daftar Pustaka:

Arhami, M., & Nasir, M. (2020). *Data Mining Algoritma Dan Implementsi*. PENERBIT ANDI. https://www.google.co.id/books/edition/Data_Mining_Algoritma_dan_Implementasi/AtcCEAAAQBAJ?hl=id&gbpv=0

DIng, W., Zhang, Y., Sun, Y., & Qin, T. (2021). An Improved SFLA-Kmeans Algorithm Based on Approximate Backbone and Its Application in Retinal Fundus Image. *IEEE Access*, *9*, 72259–72268. https://doi.org/10.1109/ACCESS.2021.3079119

Febrinita, F., Puspitasari, W., & Zaman, W. (2023). Klasterisasi Hasil Belajar Matematika dengan Algoritma K-Means Clustering. *Generation Journal*, *7*(2), 116–125. https://doi.org/10.29407/gj.v7i2.20359

Fhadli, M., & Tempola, F. (2020). *Data Mining Dengan Python Untuk Pemula*. SPASI MEDIA. https://www.google.co.id/books/edition/Data_Mining_Dengan_Python_Untuk_Pemula/8Jf_DwAAQBAJ?hl=id&gbpv=0

Jollyta, D., Siddik, M., Mawengkang, H., & Efendi, S. (2021). *Teknik Evaluasi Cluster Solusi Menggunakan Python Dan Rapidminer*. Deepublish. https://www.google.co.id/books/edition/Teknik_Evaluasi_Cluster_Solusi_Menggunak/3rcgEAAAQBAJ?hl=id&gbpv=0

Koto, S. Z., Marsono, & Syahra, Y. (2020). Analisa Data Mining Untuk Pengelompokkan Pemegang Sertipikat Hak Atas Tanah dengan Algoritma K-Means Clustering Di Kota Medan. *Jurnal CyberTech*, *x. No.x*(x).

Merly, P., Karina, D., Santi, I. H., & Puspitasari, W. D. (2024). *Pengelompokan Data Status Pertanahan Letter C menggunakan Algoritma Partitioning Around Medoids Cluster ing Land Status Data in Letter C Using the Partitioning Around Medoids Algorithm*. *12*(3), 514–521. https://doi.org/10.26418/justin.v12i3.79285

Mulaab. (2021). *Data Mining : Konsep dan Aplikasi*. Media Nusa Creative. https://www.google.co.id/books/edition/Data_Mining_Konsep_dan_Aplikasi/X1FKEAAAQBAJ?hl=id&gbpv=0

Mustika, Ardila, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Guntoro, Manuhutu, M. A., Ridwan, M., Hozairi, Wardhani, A. K., Alim, S., Romli, I., Religia, Y., Octafia, D. T., Sufandi, U. U., & Ernawati, I. (2021). *Data Mining Dan Aplikasinya*. Widina Bhakti Persada Bandung. https://www.google.co.id/books/edition/DATA_MINING_DAN_APLIKASINYA/53FXEAAAQBAJ?hl=id&gbpv=0

Nasution, A. L., & Fatonah, R. N. S. (2023). *Klasifikasi Kondisi Peralatan Elektronik Metode Gaussian Naive Bayes* (R. Habibi (ed.)). Penerbit Buku Pedia. https://www.google.co.id/books/edition/Klasifikasi_Kondisi_Peralatan_Elektronik/_nLJEAAAQBAJ?hl=id&gbpv=0

Novia, E. A., Rahayu, W. I., & Prianto, C. (2020). *Sistem Perbandingan Algoritma K-Means Dan Naïve Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan*. Kreatif. https://www.google.co.id/books/edition/SISTEM_PERBANDINGAN_ALGORITMA_K_MEANS_DA/MND9DwAAQBAJ?hl=id&gbpv=0&kptab=overview

Santi, I. H., Febrinita, F., & Puspitasari, W. D. (2023). *Engineering Design Business Process Modelling Letter C Land Data Archiving System with Software Requirement Specifications Approach*. *6*(4), 231–240.

Setiawan, E., Santi, I. H., & Budiman, S. N. (2022). Sistem Pengelolaan Dan Pengamanan Arsip Data Letter C Desa. *Jurnal Mahasiswa Teknik Informatika)*, *6*(2), 655–666.

Soepadi, H., & Widodo, P. H. (2021). Perancangan Sistem Informasi Pertanahan Buku C Desa. *IC-Tech*, *16*(1), 1–11. https://ejournal.stmik-

wp.ac.id/index.php/ictech/article/view/150

Sulianta, F. (2023). *Basic Data Mining From A to Z*. Feri Sulianta. https://www.google.co.id/books/edition/Basic_Data_Mining_from_A_to_Z/JcLhEAAAQBAJ?hl=id&gbpv=0

Sumarauw, S. J. A. (2022). *Data Mining Model Self-Organizing Maps (SOMs)*. CV. Bintang Semesta Media. https://www.google.co.id/books/edition/Data_Mining_Model_Self_Organizing_Maps_S/nI6tEAAAQBAJ?hl=id&gbpv=0

Supriyadi, A., Triayudi, A., & Sholihati, I. D. (2021). Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *6*(2), 229–240. https://doi.org/10.29100/jipi.v6i2.2008

Yuliana Sari, R., Oktavianto, H., & Wahyu Sulistyo, H. (2022). Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia. *Jurnal Smart Teknologi*, *3*(2), 104–108.