

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 *Letter – C*

Letter – C merupakan sebuah dokumen yang dapat menunjukkan hak kepemilikan tanah atau penguasaan atas suatu lahan. Dokumen *Letter – C* adalah surat minim yang memperlihatkan data sebuah properti, yang terdapat pada Kantor Desa atau Kelurahan (Lintang & Andari, 2023). Hak kepemilikan bisa berfungsi sebagai hak utama yang menjadi dasar untuk hak-hak lainnya, hak yang dimaksud adalah pemilik tanah dapat memberikan tanah kepada pihak lainnya dengan hak menyewa, menggadaikan, bagi hasil (Sugeng, 2021). *Letter – C* menjadi dokumen yang memiliki peran penting dalam membuktikan dan melindungi hak kepemilikan atas tanah, dan juga sebagai dasar untuk melakukan perubahan data tanah dalam proses transaksi dan administrasi. Dokumen ini merupakan arsip desa yang tidak dapat diganti maupun diperbaharui jika terjadi kerusakan atau kehilangan data. Gambar 2.1 adalah buku register C versi kesatu yang sudah mengalami kerusakan.



Gambar 2. 1 Buku Register C versi kesatu

Pada buku Register C atau *Letter – C* terdapat *form* nama, dan nomor register pada bagian atas, kemudian terdapat beberapa kolom dibawahnya yang mencantumkan tipe tanah sawah atau tanah kering yang masing – masing terdapat data Nomor Persil, Kelas Desa, Luas Milik, dan Sebab terjadinya perubahan seperti pada gambar 2.2. Pada bagian atas diisi dengan nama kepemilikan tanah dan juga nomor register, sedangkan pada bagian kolom Nomor Persil diisi nomor persil tanah, pada kelas desa diisi jenis atau tipe tanah, dimana daratan atau tanah kering diisi dengan menggunakan kode “d” yang memiliki 4 kategori yaitu dI sampai dengan dIV, sedangkan untuk tanah sawah dituliskan dengan kode “s” yang memiliki 3 kategori yaitu sI sampai dengan sIII. Kemudian pada kolom luas milik diisi luas tanah dengan satuan ha atau da, tetapi didalam buku register C yang dituliskan oleh pemerintahan desa Pandanarum menggunakan satuan m². Pada kolom terakhir terdapat sebab dan tanggal perobahan yang diisi histori terjadinya perubahan kepemilikan tanah. Gambar 2.2 berikut adalah form Buku Register C.

NAMA <u>WARSIH</u> No. <u>3765</u> TEMPAT TINGGAL <u>SEITUL</u>												
SAWAH						TANAH KERING						
Nomor persil dan huruf bagian persil	Kelas desa	Menurut daftar perincian				Sebab dan tanggal perobahan	Nomor persil dan huruf bagian persil	Kelas desa	Menurut daftar perincian			
		Luas milik		Ipeda					Luas milik		Ipeda	
		ha	da	R	S				ha	da	R	S
81	sI	00677				dI. no: 1710						
62	sII	00813				ket: warisan						

Gambar 2. 2 Form Buku Register C

2.1.2 *Cross-Industry Standard Process for Data Mining (CRISP – DM)*

CRISP – DM merupakan sebuah metode kerangka kerja atau metodologi yang digunakan dalam *data mining* serta analisis data (Sulianta, 2023). CRISP – DM digunakan dalam beberapa proyek *data mining* dan analisis data yang berfungsi untuk mengidentifikasi pola serta hubungan dalam data. 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. Perputarannya itu dapat berfungsi untuk memberikan pemahaman baru yang mungkin pada beberapa tahapan sebelumnya dapat dikerjakan dengan lebih baik (Sumarauw, 2022). Metode CRISP – DM merupakan metode yang sering digunakan, metode ini juga sangat terstruktur dengan terdiri dari 6 tahapan, sebagai berikut (Sulianta, 2023):

a. *Business Understanding*

Pemahaman Bisnis merupakan pemahaman mendalam tentang masalah yang ingin dipecahkan menggunakan *data mining*. Tahap ini bertujuan untuk mendefinisikan tujuan bisnis, target, serta kendala yang relevan.

b. *Data Understanding*

Merupakan tahap mengumpulkan serta memahami data yang relevan untuk proyek data mining. Yang mencakup tentang eksploitasi sumber data, eksplorasi data awal, dan pemahaman karakteristik data.

c. *Data Preparation*

Persiapan data merupakan tahap pembersihan, transformasi, serta pemodelan data untuk digunakan pada proses *data mining*. Tahap ini juga termasuk proses penghapusan data yang tidak relevan, penggabungan sumber data dan pelabelan data.

d. *Modeling*

Pemodelan data adalah tahap untuk memilih teknik *data mining* yang sesuai serta mengembangkan model analisis data. Tahap ini dapat melibatkan teknik seperti regresi, klasifikasi, klasterisasi, maupun asosiasi.

e. *Evaluation*

Tahap evaluasi merupakan tahap pengujian untuk memastikan hal tersebut memenuhi tujuan bisnis dan memiliki kinerja yang baik. Evaluasi ini melibatkan penggunaan teknik statistika serta teknik validasi untuk mengukur kualitas model.

f. *Deployment*

Pada tahap ini merupakan tahap akhir untuk mengimplementasikan model dalam lingkungan yang sebenarnya. Tahap ini dapat mencakup pengkodean model, pengintegrasian kedalam sistem serta pelatihan personel yang terlibat (Sulianta, 2023).

2.1.3 Data Mining

Data mining adalah teknologi yang digunakan untuk mengeksplorasi informasi serta pola yang berguna dari kumpulan data berukuran besar. Kegiatan data mining meliputi pengumpulan data, ekstraksi informasi, dan analisis statistik data (Arhami & Nasir, 2020). Data mining memiliki tujuan untuk menemukan

hubungan yang dapat dijadikan dalam pengambilan sebuah keputusan yang lebih baik, menentukan sebuah prediksi hasil dimasa depan, atau menyederhanakan pemahaman tentang sebuah data. Data mining dapat dijadikan alat untuk melakukan pengelompokkan data yang memiliki kesamaan karakteristik ke dalam suatu kelompok atau klaster. Dengan menggunakan *data mining* untuk menemukan data dalam jumlah yang besar melalui pengelolaan data yang tepat serta efektif menggunakan beberapa metode yang ada. Sehingga secara singkat *data mining* dapat digambarkan sebagai suatu pola, model atau kaidah pengetahuan dari *data mining* (Arhami & Nasir, 2020), seperti pada gambar 2.3



(Sumber : Arhami & Nasir, 2020)

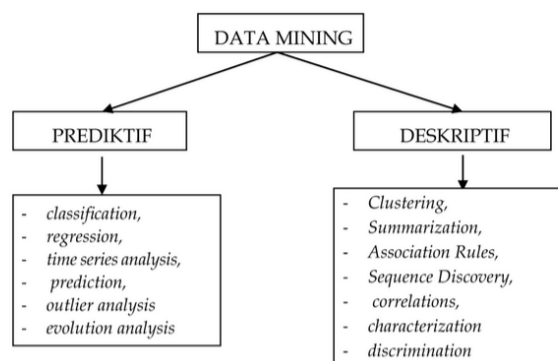
Gambar 2. 3 Model atau Pengetahuan *Output* Data Mining

Pada dasarnya *data mining* memiliki 7 fungsi dasar diantaranya Prediksi (*Prediction*), Sekuensi (*sequencing*), Klasifikasi (*Classification*), Asosiasi (*Association*), Pengelompokkan (*Clustering*), Perkiraan (*Forecasting*), dan Deskripsi (*Description*). Serangkaian proses *data mining* digunakan untuk penggalian data yang belum ditemukan sebelumnya. Proses maupun tahapan penggalian data dimulai dengan data mentah hingga berakhir pada pengetahuan atau informasi yang telah diproses (Mustika dkk., 2021). Data mining dapat dikerjakan dengan tahap - tahap berikut :

- a. Eksplorasi : mempersiapkan data dengan jumlah besar, kemudian data dibersihkan, dan menghilangkan data duplikasi untuk menghasilkan data yang dapat digunakan.

- b. Pemodelan atau identifikasi : membangun model statistik untuk mengevaluasi sehingga dapat menghasilkan nama yang baik dan akurat.
- c. Penerapan : dilakukan pengujian model terhadap data latih dan uji agar menghasilkan prediksi atau perkiraan hasil sesuai yang diharapkan (Arhami & Nasir, 2020).

Data mining memiliki 2 strategi yang digunakan untuk menyelesaikan permasalahan, yaitu Strategi Prediktif dan Strategi Deskriptif. Strategi Prediktif dimana proses sebuah model yang memiliki beberapa variabel yang digunakan untuk prediksi hasil dimasa yang akan datang. Sedangkan Strategi Deskriptif menghasilkan korelasi, tabulasi silang, frekuensi dan sebagainya, yang terfokuskan pada peringkasan serta konversi data menjadi informasi yang bermakna digunakan sebagai pelaporan dan pemantauan, seperti pada gambar 2.4. Dimana menggambarkan titik berat utama *data mining* untuk menyelesaikan masalah(Arhami & Nasir, 2020) .



(Sumber : Arhami & Nasir, 2020)

Gambar 2. 4 Strategi Data Mining

2.1.4 *Clustering Partitional*

Clustering atau pengelompokkan merupakan sebuah teknik data mining yang digunakan untuk melakukan pegelompokkan data yang serupa kedalam

kelompok atau klaster berdasarkan kesamaan dalam atribut tertentu (Setiawan dkk., 2023). Tujuan *Clustering* adalah untuk menemukan sebuah struktur yang tersembunyi didalam data tidak terstruktur atau semi terstruktur. Dengan kemampuan yang dimiliki, peran *clustering* menjadi sangat penting dimana *clustering* mampu mengungkap sebuah pola tersembunyi serta meyediakan sebuah wawasan.

Clustering memiliki beberapa jenis, dimana pada setiap jenis memiliki keunggulan serta keterbatasan masing – masing. Untuk memilih metode pengelompokkan yang tepat dapat dilihat dari jenis data, tujuan analisis, dan sumber data yang telah tersedia. Jenis *clustering* yang utama meliputi *Clustering Partitional*, *Clustering Hierarkis*, *Clustering Berbasis Densitas*, *Clustering Berbasis Model*, dan *Clustering Berbasis Grid* (Rahayu dkk., 2024).

Dalam penelitian ini pengelompokan data akan dilakukan dengan metode *Clustering Partitional*. Dalam metode tersebut akan dilakukan pengelompokkan data menjadi beberapa *cluster* tanpa ada tumpang tindih antar *cluster*. Dalam metode ini algoritma yang paling populer adalah *K – Means*, yang bekerja dengan menentukan *centroid* secara acak yang selanjutnya mengelompokkan objek berdasarkan kedekatan dengan centroid tersebut. *K – Means* efektif digunakan untuk dataset besar, tetapi sensitif pada pemilihan awal *centroid* serta cenderung terpengaruh oleh outlier. Selain *K – Means* terdapat algoritma lainnya termasuk *K – Medoids* dan *K – Medians* yang serupa dengan *K – Means*.

2.1.5 Algoritma *K – Means*

Algoritma ini merupakan salah satu algoritma dalam metode *clustering partitional* yang berfungsi untuk melakukan mengelompokkan objek data kedalam *cluster* atau kelompok – kelompok berdasarkan kesamaan fitur. Algoritma *K – Means* mengelompokkan data didalam sejumlah kelompok tertentu berdasarkan nilai rata-rata (*mean*) dimana jumlah *cluster* telah ditentukan sebelumnya. Algoritma *K – Means* sangat populer karena algoritma ini mudah diimplementasikan serta kemampuannya untuk mengklasterisasi data dalam jumlah besar dan deteksi outlier dengan sangat cepat (Nurhayati, 2022).

Algoritma *K – Means* adalah operasi pengelompokkan pada set data tanpa label serta merupakan algoritma pembelajaran tanpa pengawasan (DIng dkk., 2021). Algoritma *K – Means* melibatkan 2 proses yaitu, proses menentukan lokasi pada pusat *cluster* serta proses pencarian pada anggota setiap *cluster*. Selanjutnya dilakukan perhitungan jarak antara setiap titik data dengan pusat *cluster*. Untuk menghitung jarak data ke- i (x_i) di pusat *cluster* ke- k (c_k) diberikan nama (d_{ik}), yang dapat dilakukan dengan menggunakan persamaan *Euclidean*. Sebuah data akan tergabung dalam *cluster* ke- k jika jarak titik data tersebut dengan titik pusat *cluster* ke- k memiliki nilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* yang lainnya (Nurhayati, 2022).

Menurut (Febrinita dkk., 2023) berikut adalah langkah – langkah pengelompokan algoritma *K – Means Clustering* :

- a. Menentukan jumlah *cluster*

Pada penelitian ini jumlah *cluster* ditentukan dengan perhitungan metode *elbow*.

- b. Memilih pusat *cluster* secara *random*
- c. Melakukan perhitungan jarak antar data dan titik pusat *cluster* dengan menggunakan persamaan *Euclidean Distance* seperti pada persamaan 2.1 berikut.

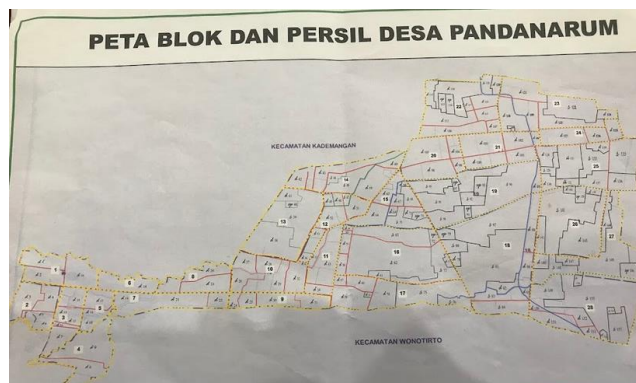
$$d(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \dots\dots\dots(2.1)$$

dimana :

$d(i, j)$ = Jarak data ke i ke pusat *cluster* j
 x_{ki} = Data ke i pada atribut data ke k
 x_{kj} = Titik pusat ke j pada atribut ke k

- d. Menentukan kembali titik pusat *cluster* baru dari nilai rata – rata setiap *cluster*.
- e. Melakukan perhitungan kembali jarak antar data dengan titik pusat *cluster* baru.
- f. Perulangan untuk menentukan pusat *cluster* serta penempatan data pada *cluster* sampai tidak terdapat perubahan pada titik pusat *cluster* (Febrinita dkk., 2023).

Pada penelitian yang akan dilakukan peneliti menggunakan 3 atribut yaitu, Nomor Blok, Nomor Persil, dan Luas milik. Ketiga atribut tersebut diperoleh dari buku Register C dan peta tanah yang terdapat di Desa Pandanarum Kecamatan Sutojayan. Di desa Pandanarum ini memiliki 28 Blok yang dapat dilihat dari peta tanah desa Pandanarum seperti gambar 2.5 berikut.



Gambar 2. 5 Peta Desa Pandanarum

Di Desa Pandanarum memiliki jumlah persil sebanyak 157 persil dengan 2 tipe kelas yaitu, daratan yang disimbolkan dengan “d” dan perairan atau persawahan yang disimbolkan dengan “s”. Kemudian untuk pembagian *cluster* akan ditentukan menggunakan *elbow method* untuk mendapatkan jumlah *cluster* terbaik. Pengelompokkan ini berdasarkan data yang diambil dari buku Register C yang ada di Desa Pandanarum, seperti gambar 2.6 berikut.

NAMA <u>WARSIH</u> No. <u>3765</u> TEMP							
SAWAH							
Nomor persil dan huruf bagian persil	Kelas desa	Menurut daftar perincian				Sebab dan tanggal perobahan	Nomor persil dan huruf bagian persil
		Luas milik		Ipeda			
		ha	da	R	S		
81	S ^I	006	77			dr. no: 1710	
62	S ^{II}	008	13			ket: warisan	

Gambar 2. 6 Form Register C Yang Terdapat Kelas Desa

2.1.6 Normalisasi Data

Normalisasi data merupakan proses pengaturan nilai dalam suatu dataset menjadi rentang yang lebih terstandarisasi atau normal. Normalisasi bertujuan untuk melakukan keseimbangan data dalam dataset, sehingga memudahkan analisa serta pemrosesan lebih lanjut. Normalisasi dilakukan sebelum proses perhitungan jarak sebuah data (Arhami & Nasir, 2020). Proses normalisasi dilakukan dimana

nilai – nilai dalam data dibagi dengan nilai data maksimum dan nilai data minimum dari data tersebut. Normalisasi data dilakukan menggunakan persamaan sebagai berikut (Mulaab, 2021):

$$f(x) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \dots\dots\dots(2.2)$$

dimana :

X_i = data ke – i
 $X_{\min}$ = nilai minimum data
 $X_{\max}$ = nilai maksimum data

2.1.7 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 akan dilakukan hingga mendapatkan nilai pengelompokan yang konvergen (Jollyta dkk., 2021). Persamaan *Euclidean Distance* yang dimaksud sama dengan persamaan 2.1

2.1.8 *Elbow Method*

Metode *elbow* adalah sebuah teknik optimalisasi untuk menentukan jumlah *cluster*, yang dihasilkan 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 akan dijadikan sebagai sumber informasi (Jollyta dkk., 2021). Tujuan dari metode *elbow* ini adalah untuk menemukan titik dimana penambahan jumlah *cluster* tidak lagi memberikan peningkatan yang signifikan dalam pengurangan varian data. Apabila nilai *cluster* ke-satu dan *cluster* ke-dua menghasilkan sudut pada *grafik* atau menunjukkan pengurangan paling signifikan maka nilai *cluster* tersebut yang tepat.

Sum of Square Error (SSE) merupakan persamaan untuk memperoleh perbandingan dari setiap nilai *cluster*. Jika titik siku yang dibentuk jaraknya semakin jauh, maka jumlah *cluster* akan menjadi lebih optimal atau paling optimal. Adapun langkah – langkah metode *elbow* sebagai berikut (Yuliana Sari dkk., 2022):

- a. Menentukan nilai awal k
- b. Menaikkan nilai k
- c. Melakukan perhitungan hasil *Sum of Square Error* dari setiap *cluster* menggunakan persamaan 2.3 berikut :

$$SEE = \sum_{K=1}^K \sum_{x_i} \|x_{ij} - c_k\|_2^2 \dots\dots\dots(2.3)$$

dimana :

K = jumlah *cluster*

x_i = jumlah titik data dalam *cluster* ke – i

x_{ij} = data ke -j dalam *cluster* ke – i

C_i = pusat *cluster* ke – i

- d. Melakukan perhitungan SSE hingga *cluster* yang ditentukan
- e. Melihat hasil SSE dari nilai *cluster* yang mengalami penurunan secara signifikan.
- f. Menetapkan nilai untuk *cluster* yang berbentuk siku

2.1.9 *Davies Bouldin Index*

Pengujian atau evaluasi kinerja dari algoritma *K-Means Clustering* berfungsi untuk mengukur kemiripan dari setiap sampel pada setiap kelas label. Pada penelitian yang akan dilakukan ini pengujian atau evaluasi yang digunakan adalah metode *Davies Bouldin Index* (DBI), metode ini bekerja dengan mempertimbangkan kesamaan rata – rata antara setiap *cluster* dan mencari *cluster* yang serupa (Fhadli & Tempola, 2020). Pengukuran DBI bertujuan untuk

memaksimalkan jarak antar sampel dalam satu kelompok atau kelas, sehingga DBI digunakan untuk menguji validasi data dalam setiap kelompok (Mustika dkk., 2021). Berikut adalah langkah – langkah perhitungan DBI :

- a. Menghitung jarak *intra-cluster* untuk standar deviasi antar titik data didalam *cluster* dengan pusat *cluster*.

$$s_i = \sqrt{\frac{1}{T_i} \sum_{j=1}^{T_i} |X_j - A_i| \cdot |X_j - A_i|} \dots\dots\dots(2.4)$$

dimana :

T_i = banyak data pada *cluster* i

X_j = vektor pada *cluster* i

A_i = pusat *cluster* i

- b. Menghitung jarak *inter-cluster* untuk jarak setipa pasangan *cluster*

$$R_{ij} = \sqrt{\|a_i - a_j\| \cdot \|a_i - a_j\|} \dots\dots\dots(2.5)$$

dimana :

a_i = Pusat *cluster* i

a_j = Pusat *cluster* j

$i \neq j$

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

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \dots\dots\dots(2.6)$$

dimana :

S_i = jumlah *cluster* i

S_j = jumlah *cluster* j

M_{ij} = jarak antara *cluster* i dan j

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

$$R_i = \max R_{ij} \dots\dots\dots(2.7)$$

e. Menghitung DBI sebagai rata – rata ukuran kesamaan maksimal *cluster*

$$DBI = \frac{1}{K} \sum_{i=1}^K R_i \dots \dots \dots (2.8)$$

dimana :

K = banyaknya *cluster*

Dari hasil nilai yang diperoleh setelah melakukan proses perhitungan dapat disimpulkan, semakin kecil nilai DBI mendekati 0 dan tidak bernilai negatif maka jumlah *cluster* yang terbentuk akan semakin baik (Fhadli & Tempola, 2020).

2.1.10 Phyton

Pemrograman bahasa phyton dikembangkan oleh Guido van Rossum pada tahun 1991 yang merupakan bahasa pemrograman bertingkat tinggi. Pemrograman phyton dikembangkan dengan kode dan sintak sederhana yang membuat pemrograman ini menjadi pilihan populer bagi pemula maupun pengembang aplikasi berpengalaman (Maesaroh dkk., 2024). Bahasa pemrograman ini memiliki beberapa fitur utama diantaranya :

1. Menggunakan bahasa kode yang sederhana sehingga mudah dibaca dan dipahami.
2. Dapat dijalankan dalam berbagai OS (Windows, macOS, dan Linux).
3. Interpretasi Kode phyton akan dieksekusi baris per baris oleh interpreter Phyton.
4. Dukungan untuk berbagai paradigma pemrograman

Bahasa ini mendukung pemrograman berorientasi objek, pemrograman fungsional, dan pemrograman berbasis aturan.

5. Bahasa ini tidak perlu mendeklarasikan tipe data secara eksplisit saat membuat variabel.

6. Memiliki berbagai library dan framework

Python memiliki beberapa *framework* seperti Django untuk pengembangan web, NumPy untuk komputasi numerik, serta TensorFlow untuk kecerdasan buatan.

7. Dapat diperluas dalam pemrograman bahasa lain seperti C atau C++.

8. Komunitas yang besar

Python memiliki komunitas yang besar dan aktif, sehingga mudah menemukan dukungan dan tutorial.

2.2 Kajian Penelitian

Dari penelitian - penelitian sebelumnya yang melakukan penelitian tentang Algoritma *K – Means Clustering* dan Digitalisasi data *Letter – C* atau buku Register C Pertanahan Desa diantaranya :

Penelitian yang dilakukan oleh (Santi dkk., 2023) yang berjudul *Engineering Design Business Process Modelling Letter C Land Data Archiving System with Software Requirement Specifications Approach* bertujuan untuk merancang sebuah model bisnis yang digunakan sebagai proses pengumpulan data serta pencarian data *Letter – C* pertanahan desa. Penelitian ini menghasilkan sebuah identifikasi yang melakukan identifikasi analisa terkait kebutuhan dari sistem baik secara fungsional maupun tidak fungsional serta sebuah perancangan sistem. Dimana dari sistem tersebut diharapkan dapat menjadi sebuah rujukan yang nantinya akan dipergunakan untuk melakukan penelitian selanjutnya.

Penelitian yang dilakukan oleh (Setiawan dkk., 2022) bertujuan untuk mengelola dan mengamankan data arsip dengan basis web agar efektif dan efisien. Dengan hasil pengujian *blackbox* testing menunjukkan semua aplikasi berfungsi

dengan baik. Sistem aplikasi ini dikembangkan menggunakan *System Development Life Cycle* (SDLC) dengan model *Prototype* serta mengembangkan sistem dengan pemrograman PHP serta *database MySQL*.

Penelitian yang dilakukan oleh (Rahmawati dkk., 2021) yang berjudul Digitalisasi Registrasi Pertanahan Kelurahan Karangbesuki Melalui Aplikasi Buku Registrasi Kelurahan (*Letter C*) Berbasis *Website* ini bertujuan untuk membuat sebuah sistem informasi pengarsipan data *letter – C* berbasis *website*. Pada penelitian ini menghasilkan sebuah aplikasi yang dapat berjalan dengan baik serta sudah dilaksanakan pengujian serta pelatihan kepada administrator.

Penelitian yang dilakukan oleh (Lintang & Andari, 2023) dengan judul Keabsahan Letter C Sebagai Alat Bukti Kepemilikan Tanah, bertujuan untuk meneliti keabsahan dokumen *Letter – C* merupakan lembar bukti kepemilikan tanah. Dengan hasil yang didapatkan adalah, *Letter – C* desa merupakan surat kepemilikan tanah yang didalamnya mencatat pemilik hak atas tanah di suatu wilayah. *Letter – C* adalah surat yang turun temurun yang dapat dijadikan sebagai perlindungan hukum bagi pemegang bukti dan dapat digunakan untuk melakukan pendaftaran tanah.

Penelitian yang dilakukan oleh (Fiokta Wiharifin, 2022) yang berjudul Sistem Rekapitulasi Data *Letter – C* Desa Berbasis *Website*, memiliki tujuan untuk membuat sebuah Sistem yang digunakan untuk rekap Data *Letter – C* berbasis *website* menggunakan metode RAD. Penelitian menghasilkan sebuah Sistem Informasi Data *Letter – C* dengan harapan dapat mempermudah staff pemerintahan desa dalam pencarian data kepemilikan tanah dan menambahkan fitur *generate*

data, yang digunakan sebagai pengantar surat tanah dari desa. Serta membantu proses pendataan staff pemerintahan desa.

Penelitian yang dilakukan oleh (Rohadi, 2021) yang berjudul Digitalisasi Registrasi Desa (*Letter C*) Tanah dalam Optimalisasi Pelayanan ditengah Pandemi *Covid19* di Pemerintah Desa Kecamatan Susukan Kabupaten Cirebon, bertujuan untuk meningkatkan pelayanan *letter c* desa dengan memanfaatkan kemajuan teknologi dengan tujuan mendigitalisasi agar terciptanya pemerintahan yang baik (*good governance*). Penelitian ini menghasilkan sebuah sistem aplikasi berbasis web yang berfungsi sebagai arsip digital pemerintahan desa dalam bidang pertanahan yang terfokus ke data *letter – C*, dengan menggabungkan *letter – C* desa dengan DHKP (SPPT) sehingga setiap data kepemilikan tanah diarsipkan dengan baik.

Jurnal penelitian terdahulu terkait data *Letter – C* terdapat pada tabel 2.1 berikut.

Tabel 2. 1 Jurnal Penelitian Terdahulu Terkait *Letter - C*

Judul	Penulis, Tahun,Jurnal	Tujuan	Hasil	Kontribusi Penelitian
<i>Engineering Design Business Process Modelling Letter C Land Data Archiving System with Software Requirement Specifications Approach</i>	Indyah Hartami Santi, Filda Febrinita, Wahyu Dwi Puspitasari, Desember 2023, <i>Journal of Intelligent Decision Support System (IDSS)</i>	Tujuan dari penelitian ini adalah merancang model bisnis untuk proses pengumpulan data serta pencarian data <i>Letter – C</i> pertanahan.	Mengidentifikasi kasi dan analisis kebutuhan sistem, merancang sistem.	Pada penelitian ini peneliti melakukan digitalisasi <i>letter – C</i> di Desa Pandanarum guna untuk meningkatkan pelayanan di pemerintahan Desa Pandanarum. Penelitian relate dengan apa yang akan dilakukan oleh penulis terkait data <i>Letter – C</i> di Desa Pandanarum

Judul	Penulis, Tahun,Jurnal	Tujuan	Hasil	Kontribusi Penelitian
Sistem Pengelolaan Dan Pengamanan Arsip Data Letter C Desa (Studi Kasus : Kantor Desa Gondang)	Eri Setiawan, Indyah Hartami Santi, Saiful Nur Budiman (September 2022), JATI (Jurnal Mahasiswa Teknik Informatika)	Untuk mengelola dan mengamankan data arsip dengan basis web agar efektif dan efisien.	Aplikasi pengelolaan sera pengamanan data arsip berbasis web, evaluasi dengan pengujian <i>black box testing</i> , kuesioner pengguna serta validasi ahli dari IT.	Penelitian tentang data pertanahan tingkat Desa atau data <i>Letter C</i> yang dapat membantu penelitian yang akan dilakukan penulis untuk mengelompokkan Data Kelas Desa yang terdapat didalam Data <i>Letter C</i>
Digitalisasi Registrasi Pertanahan Kelurahan Karangbesuki Melalui Aplikasi Buku Registrasi Kelurahan (<i>Letter C</i>) Berbasis <i>Website</i>	Noni Mia Rahmawati, Adekun Cahyono, Salma Shofia, Amirah Tsara Nafisatul Sakinah, Desember 2021, Jurnal Warta Desa (JWD)	Penelitian yang bertujuan untuk membuat sistem informasi pengarsipan data <i>Letter-C</i> berbasis <i>web</i> .	Aplikasi buku Register C berbasis <i>website</i> , yang dilakukan dengan 3 tahap yaitu, dalam tahap persiapan, tahap pelaksanaan, serta tahap evaluasi.	Penelitian ini <i>relate</i> dengan apa yang akan peniliti lakukan saat ini , sehingga dapat menjadi referensi bagi peneliti
Keabsahan Letter C Sebagai Alat Bukti Kepemilikan Tanah	Ayu Lintang Priyan Andari, Miftahul Hasanah, Sumriyah, Februari 2023, Jurnal Hukum dan Sosial Politik	Dari penelitian ini memiliki tujuan untuk bagaimana keabsahan Letter-C sebagai bukti kepemilikan tanah.	<i>Letter – C</i> merupakan surat yang mencatat kepemilikan tanah dalam suatu wilayah secara turun menurun.	Penelitian ini dapat dijadikan refensi bagi penulis untuk penelitian yang akan dilakukan

Judul	Penulis, Tahun,Jurnal	Tujuan	Hasil	Kontribusi Penelitian
Sistem Rekapitulasi Data <i>Letter C</i> Desa Berbasis <i>Website</i>	Ressy Fiolta Wiharifin, 2023 Seminar Nasional Teknologi Informasi dan Komunikasi-2023	Membangun sebuah Sistem untuk merekap Data <i>Letter-C</i> berbasis <i>website</i> dengan menerapkan meto de RAD.	Pengembangan aplikasi yang dilakukan rekapitulasi data <i>letter-C</i> basis web, dengan metode RAD, di bangun dengan bahasa pemrograman <i>php</i> , <i>html</i> , <i>css</i> , dan <i>database mysql</i> .	Dari penelitian yang dilakukan dapat dijadikan referensi oleh penulis karena penelitian ini <i>relate</i> dengan apa yang akan dilakukan penulis yaitu melakukan pengamatan terhadap buku <i>letter c</i> desa, yang kemudian dilakukan digitalisasi.
Digitalisasi Registrasi Desa (<i>Letter C</i>) Tanah dalam Optimalisasi Pelayanan ditengah Pandemi <i>Covid19</i> di Pemerintah Desa Susukan Kecamatan Kabupaten Cirebon	Abdul Wahid, Rohadi, Desember 2021, Mahkamah: Jurnal Kajian Hukum Islam	Untuk memberikan pelayanan registrasi desa (<i>letter C</i>) yang memanfaatkan kemajuan teknologi menggunakan konsep digitalisasi sehingga tercipta pemerintahan yang baik (<i>good governance</i>)	Sebuah aplikasi yang berbasis web (<i>local host</i>), dengan penggabungan Buku Register C, DHKP (SPPT) dengan tujuan setiap data kepemilikan tanah tersiapkan dengan baik.	Pada penelitian ini membahas tentang digitaslisasi Register C atau <i>Letter C</i> , dimana ini dapat menjadi referensi bagi pelitian yang akan dilakukan, yaitu mengelompokkan data kelas desa

Penelitian yang dilakukan oleh (Febrinita dkk., 2023) yang berjudul Klasterisasi Hasil Belajar Matematika dengan Algoritma *K-Means Clustering* memiliki tujuan untuk mengklasterisasi atau mengelompokkan nilai belajar matematika mahasiswa. Penelitian ini menghasilkan 5 *cluster* dengan nilai tertinggi yaitu *cluster dua* (2) meraih nilai 86,81 sedangkan nilai rerata terendah yaitu cluster 5 memperoleh nilai 76,50. Dalam *cluster* ke-2 lebih dominan dengan siswa lulusan dari SMK yang masuk jurusan TKJ. Sedangkan di *cluster* ke-5 lebih

dominan oleh siswa dari lulusan SMA dengan jurusan IPA, dan juga terdapat temuan jika Siswa SMK tidak selalu menunjukkan kemampuan matematika yang lebih rendah dibandingkan siswa SMA, karena tingkat keberhasilan belajar dapat dipengaruhi oleh motivasi intrinsik

Penelitian yang dilakukan oleh (Koto, dkk., 2020) yang berjudul *Analisa Data Mining Untuk Pengelompokan Pemegang Sertipikat Hak Atas Tanah dengan Algoritma K-Means Clustering Di Kota Medan* ini bertujuan untuk mengelompokkan Pemegang Sertifikat Hak Atas Kepemilikan Tanah pada setiap wilayah untuk melihat wilayah di kota medan yang telah banyak dikeluarkan sertifikat tanah sehingga dapat menentukan daerah yang akan dijadikan sebagai target pengeluaran sertipikat secara besar-besaran dan gratis. Dengan hasil yang pertama yaitu, penerapan permasalahan yang terjadi terkait penerbitan sertipikat tanah menggunakan Algoritma *K – Means Clustering*, kedua analisa kelompok pemegang sertifikat dan wilayah yang kayak diterbitkan sertipikat secara massal dan gratis, serta sebuah sistem yang dapat digunakan untuk melakukan pengelompokan Pemegang Sertifikat Hak Atas Kepemilikan Tanah di Kota Medan.

Penelitian yang dilakukan oleh (Adiputra, 2022) yang berjudul *Clustering Penyakit DBD Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means*, bertujuan untuk dapat mengetahui gambaran atau perkembangan yang terjadi didalam data diagnisa penyakit DBD. Penelitian ini menghasilkan bahwa, pasien dengan diagnosa penyakit DBD adalah pasien dengan rentang umur 25 tahun sampai dengan 45 tahun, dengan estimasi jumlah hari perawatan yang tidak sama,

sehingga dapat disimpulkan bahwa usia tidak dapat mempengaruhi berapa lama pasien akan dirawat.

Penelitian yang dilakukan oleh (Fitrianti dkk., 2023) yang berjudul *Clustering Film Populer Pada Aplikasi Netflix Dengan Menggunakan Algoritma K-Means Dan Metode CRISPDM*, bertujuan untuk mengelompokkan film – film populer di aplikasi Netflix. Penelitian tersebut menghasilkan tiga (3) *cluster* dengan rata – rata centroid yang berbeda, dimana *cluster* pertama menjadi klaster dengan kategori ideal dalam mengelompokkan film populer netflix yang memiliki rata – rata rating sebesar 8.571, durasi 166.036,00 dan vote sebesar 1.157.600,464.

Penelitian yang dilakukan oleh (Supriyadi dkk., 2021) tentang Perbandingan Algoritma *K-Means* Dengan *K-Medoids* Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas, menghasilkan nilai validasi DBI sebesar 0,67 untuk *K – Means* dan 1,78 untuk *K-Medoids*, sehingga aplikasi berbasis web akan dibuat dengan mengimplementasikan Algoritma *K-Means* karena hasil nilai DBI lebih rendah dari *K-Medoids*. Hasil uji dari aplikasi berbasis web dengan mengimplemetasikan Algoritma *K-Means* memiliki kesesuaian sebesar 97%, yang diuji menggunakan *tool Rapidminer* maupun diuji dengan perhitungan secara manual.

Penelitian yang dilakukan oleh (Tanjung dkk., 2021) dengan judul Penerapan Metode *K-Means* Pada Pengelompokkan Pengangguran Di Indonesia bertujuan untuk menganalisa jumlah pengangguran yang terdapat di setiap Provinsi di Indonesia. Tujuan dari penelitian ini diharapkan bisa membantu Pemerintah

dalam menyediakan lapangan pekerjaan yang lebih luas dan membangun serta meningkatkan perekonomian setiap daerah provinsi di Indonesia.

Penelitian yang dilakukan oleh (DIng dkk., 2021) yang berjudul *An Improved SFLA-Kmeans Algorithm Based on Approximate Backbone and Its Application in Retinal Fundus Image* bertujuan untuk melakukan pengembangan dan pengujian metode baru yang dapat meningkatkan kinerja algoritma *K – Means* dengan menggabungkan algoritma *Shuffled frog leaping* untuk mendapatkan hasil klasterisasi yang lebih baik dalam waktu yang cepat dan singkat. Dari penelitian ini dihasilkan sebuah metode baru yang menggabungkan algoritma *shuffled frog leaping* dengan algoritma *K – Means* berhasil meningkatkan kinerja dalam pengelompokan menjadi lebih baik serta mengurangi waktu eksekusi.

Penelitian yang dilakukan oleh (Sammouda & El-Zaart, 2021) yang berjudul *An Optimized Approach for Prostate Image Segmentation Using K-Means Clustering Algorithm with Elbow Method* menyajikan pendekatan terotomasi untuk segmentasi gambar prostat menggunakan algoritma *K – Means* dengan metode *Elbow*. Dengan metode ini dapat menghasilkan pengelompokan dan analisis piksel gambar histologis dan gambar kanker prostat. Hasil eksperimen pada dua set data gambar prostat warna diperoleh menggunakan pendekatan segmentasi gambar yang dioptimalkan. Analisis gambar yang di-segmentasi dengan lebih banyak daya tarik pada *cluster* dengan wilayah kecil dapat membantu mencapai diagnosis yang lebih sempurna.

Penelitian terdahulu terkait metode *Elbow*, algoritma *K – Means Clustering*, metode evaluasi DBI, dan metode penelitian CRISP – DM terdapat pada tabel 2.2 berikut :

Tabel 2. 2 Jurnal Penelitian Terdahulu Pengelompokan (*K - Means*)

Judul	Penulis, Tahun,Jurnal	Tujuan	Hasil	Kontribusi Penelitian
Klasterisasi Hasil Belajar Matematika dengan Algoritma <i>K-Means Clustering</i>	Filda Febrinita, Wahyu Dwi Puspitasari, Wahid Ibnu Zaman. Juli 2023, <i>Generation Journal</i>	Mengklasterisasi hasil belajar matematika mahasiswa	Menghasilkan 5 <i>cluster</i> . Dimulai dari urutan tertinggi klaster 2 dengan nilai rata – rata 86,81, dan urutan terendah pada klaster 5 dengan nilai rata – rata 76,50.	Pada penelitian ini penerapan algoritma <i>K-Means Clustering</i> dapat menjadi rujukan yang relate dengan apa yang akan akan dilakukan oleh penulis, yang juga menggunakan algoritma <i>K-Means</i>
Analisa <i>Data Mining</i> Untuk Pengelompokan Pemegang Sertipikat Hak Atas Tanah dengan Algoritma <i>K-Means Clustering</i> Di Kota Medan	Siti Zubaedah Koto, Marsono, Yohanni Syahra, September 2020, Jurnal <i>CyberTech</i>	Mengelompok an Pemegang Sertifikat Hak Atas Kepemilikan Tanah, agar dapat menentukan wilayah yang akan dijadikan sebagai target pengeluaran sertipikat secara massal dan gratis.	Penerapan permasalahan terkait penerbitan sertifikat tanah menggunakan Algoritma <i>K – Means Clustering</i> . Cara analisa pengelompokan. Merancang aplikasi pengelompokan Pemegang Sertipikat Hak Atas Tanah di Kota Medan	Penelitian ini relate dengan penelitian yang akan dilakukan, dimana menggunakan Algoritma <i>K – Means</i> , sehingga dapat menjadi referensi untuk melakukan penelitian terkait <i>Letter – C</i> .
Clustering Penyakit DBD Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma <i>K-Means</i>	I Nyoman Mahayasa Adiputra, Desember 2021, INSERT: <i>Information System and Emerging Technology Journal</i>	Untuk memahami gambaran yang terjadi dalam data diagnosa DBD	Diagnosa pasien DBD terbagi dalam 3 klaster, dengan jumlah paling rendah pada interval usia 25 – 45 tahun. Dan usia pasien tidak mempengaruhi lama hari perawatan.	Pada penelitian ini yang dilakukan dengan penerapan algoritma <i>K-Means Clustering</i> relate dengan penelitian yang akan dilakukan
Clustering Populer Pada Aplikasi Netflix Dengan Menggunakan	Ika Fitrianti, Apriade Voutama, Yuyun Umaidah.	Untuk meng cluster atau mengelompo kan film yang populer di	Menghasilkan 3 klaster berdasarkan jumlah durasi, rating, dan <i>votes</i>	Melakukan pengelompokan terhadap film populer pada aplikasi netflix

Judul	Penulis, Tahun,Jurnal	Tujuan	Hasil	Kontribusi Penelitian
Algoritma K-Means Metode CRISPDM	K-Dan 2023, JTSTI	aplikasi Netflix	dengan kategori rendah, sedang, dan tinggi. Dari ketiga klaster ciri – ciri ideal <i>clustering</i> film populer terdapat pada klaster 1 dengan nilai rata – rata rating 8.571, nilai rata – rata durasi 166.036, dan nilai rata – rata vote 1.157.600,464.	denga menggunakan metode CRISP – DM. Hal ini relate dengan apa yang akan dilakukan peneliti, dimana peneliti juga menggunakan metode CRISP - DM
Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas	Aceng Supriyadi, Agung Triayudi, Ira Diana Sholihati Desember 2021,JIPI	Meningkatkan akurasi dan efisiensi kinerja armada truk dalam bisnis transportasi	Dari pengujian yang dilakukan nilai DBI yang dihasilkan 0,67 dari K-Means dan 1,78 dari K-Medoids, dan implementasi berbasis web dengan menggunakan hasil perhitungan K-Means.	Pada penelitian ini akan menjadi referensi dalam menentukan nilai hasil evaluasi menggunakan DBI, sehingga relate dengan pengujian yang akan dilakukan.
Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia	Fadhillah Azmi Tanjung, Agus Perdana Windarto, M Fauzan. 1 Februari, Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)	Untuk melakukan analisis tingkat pengangguran menurut Provinsi yang ada di Indonesia.	Mengelompokan data pengangguran menjadi 2 klaster berdasarkan provinsi. Yang menunjukkan 13 provinsi dengan angka tertinggi dalam klaster 1, dan 21 provinsi terendah dalam klaster 2	Dari jurnal yang dilakukan pengelompokan dengan implementasi algoritma K – Means relate dengan algoritma yang akan digunakan oleh penulis
An Improved SFLA-Kmeans Algorithm Based on Approximate Backbone and Its Application in Retinal Fundus Image	Weiping Ding, (Senior Member, Ieee), Yi Zhang, Ying Sun, And Tingzhen Qin, Mei 2021, IEEE Access	Mengembangkan an dan menguji metode baru yang dapat meningkatkan kinerja algoritma K – Means Clustering, dengan	Menunjukkan metode baru yang menggabungkan algoritma <i>shuffled frog leaping</i> dengan algoritma K – Meas berhasil meningkatkan kinerja dalam klasterisasi	Penelitian ini memiliki kesamaan dengan apa yang akan dilakukan oleh peneliti. Dengan menggunakan algoritma K – Means.

Judul	Penulis, Tahun,Jurnal	Tujuan	Hasil	Kontribusi Penelitian
		menggabungkan algoritma <i>shuffled frog leaping</i> yang terfokus pada algoritma <i>K-means</i> .	menjadi lebih baik dan mengurangi waktu eksekusi.	
<i>An Optimized Approach for Prostate Image Segmentation Using K-Means Clustering Algorithm with Elbow Method</i>	Rachid Sammouda, and Ali El-Zaart, 2021, <i>Computational Intelligence and Neuroscience</i>	Mengelompokkan dan menganalisis piksel gambar histologis dan gambar kanker prostat dengan inframerah	Dus set data yang dianalisis dengan segmentasi gambar yang optimal,	Pada penelitian ini menggunakan <i>elbow method</i> untuk menentukan <i>cluster relate</i> dengan apa yang akan dilakukan oleh penulis

Dari penelitian yang dikerjakan oleh Santi dkk., (2023) menghasilkan sebuah rancangan rekayasa sistem pengumpulan, pencarian, dan pemetaan data *Letter – C* di Desa Pandanarum. Penelitian yang diselesaikan oleh Setiawan dkk., (2022) menghasilkan sebuah sistem pengelolaan dan pengamanan data *Letter – C* desa. Penelitian yang dilakukan oleh Koto dkk., (2020) menghasilkan cara pengelompokan Pemegang Sertipikat Hak Atas tanah menggunakan *K – Means*. Dan penelitian yang dikerjakan oleh Supriyadi dkk., (2021) menunjukkan bahwa berdasarkan nilai kesamaan algoritma *K – Means* lebih tepat untuk diterapkan pada pembangunan aplikasi Pengelompokan Aramada Kendaraan berbasis web dengan nilai DBI lebih rendah dibandingkan dengan *K – Medoids*. Dari jurnal penelitian diatas relate dengan penelitian yang akan dilakukan. Pada pelitian diatas menghasilkan sebuah sistem aplikasi berbasis web dan pengelompokan menggunakan *K – Means*. Sedangkan pada penelitian yang akan dilakukan adalah terkait digitalisasi dokumen *letter – C* dan pengelompokan data kelas desa sesuai dengan kategori.