

BAB II

TINJAUAN PUSTAKA

2.1. Landasan Teori

2.1.1. Pajak Bumi dan Bangunan

Berdasarkan Undang-undang Nomor 12 Tahun 1985 tentang Pajak Bumi dan Bangunan sebagaimana telah diubah dengan Undang-undang Nomor 12 Tahun 1994, Pajak Bumi dan Bangunan (PBB) adalah pajak negara yang dikenakan setiap tahun. PBB adalah pajak kebendaan, yang berarti besarnya pajak yang terutang ditentukan oleh keadaan objek, yaitu tanah dan struktur. Besar pajak tidak dipengaruhi oleh keadaan pembayar. PBB kemudian mengubah wilayah pedesaan dan perkotaan menjadi Pajak Daerah, yang diatur dalam Undang-Undang No.28 Tahun 2009 tentang Pajak Daerah dan Retribusi Daerah (PDRD) Pasal 77 hingga 84 mulai tahun 2010. Subjek pajak Pajak Bumi dan Bangunan sekaligus menjadi wajib pajak. Subjek pajak dalam Pajak Bumi dan Bangunan adalah individu atau badan yang secara nyata memiliki hak atas tanah, atau memiliki, menguasai, atau memperoleh manfaat atas tanah tersebut (Syarifudin, 2018).

Menurut Bab I UU No. 28 Tahun 2009 berisi Ketentuan Umum yang menjelaskan istilah-istilah teknis dan definisi PBB:

1. Tubuh bumi dan permukaannya adalah bumi. Pengertian ini mencakup semua bagian tubuh bumi, mulai dari permukaan hingga magma, produk tambang, dan gas material lainnya.
2. Bangunan adalah struktur teknik yang ditanam atau diikat pada tanah dan/atau air.

Adapun dokumen ketetapan pajak yakni Dokumen Daftar Himpunan Ketetapan Pajak DHKP yang merupakan dokumen Dokumen yang berisi kumpulan ketetapan dan pembayaran pajak, berisi pencatatan setiap rumah dan pemilik tanah untuk mengatur wajib pajak yang berikan pada setiap pemilik tanah. Pendataan buku ini dilakukan secara setiap tahun. selain itu ada dokumen tanah yakni Buku Desa C atau *Letter C* merupakan buku yang digunakan oleh masyarakat sejak zaman kolonial Belanda hingga sebelum tahun 1960 an. Saat ini buku tersebut digunakan sebagai bukti hak milik atas tanah yang belum berstatus bersertifikat. Karena tanah pada Buku C sudah bertahun-tahun dibuat secara tertulis, maka atas dasar itu notaris dan agen pertanahan dapat melihat siapa pemilik hak atas tanah yang belum bersertifikat tersebut. Dengan adanya peraturan pemerintah No. 24 Tahun 1997 dengan jelas mengatur pengertian pendaftaran tanah, khususnya Pasal 1 ayat (1), secara tegas, "Pendaftaran tanah adalah suatu rangkaian kegiatan yang dilaksanakan oleh pemerintah secara terus-menerus, terus-menerus dan teratur, meliputi pengumpulan, pengolahan, pencatatan, presentasi dan pemeliharaan data fisik dan hukum berupa peta dan daftar bidang" (Astuti, 2022).

Dapat dijelaskan *Letter C* menjadi dokumen penting di setiap desa karena memuat catatan sejarah dan sejarah tanah yang datanya harus dilestarikan. Desa ini memiliki beberapa arsip penting yang tidak dapat diperbarui atau diganti jika rusak atau hilang. Salah satu arsip penting yang dimiliki oleh desa adalah dokumen yang berhuruf C. Kantor desa menyimpan banyak dokumen rahasia, termasuk data kependudukan dan data lainnya. Masalah keamanan dan kerahasiaan data sangat penting di kantor desa (Setiawan dkk., 2022).

2.1.2. Pajak

Pajak, yang merupakan transfer kekayaan dari sektor swasta ke negara, merupakan kontribusi yang diatur secara hukum dari warga negara ke kas negara. Mereka tidak ditunjuk secara langsung dan tidak menerima jasa pimpinan yang dapat digunakan untuk mencegah atau memperlancar pembiayaan pengeluaran umum atau pencapaian tujuan non-keuangan. Pajak memegang peranan yang sangat penting dalam kehidupan suatu bangsa, khususnya dalam pelaksanaan pembangunan. Sebab, pajak merupakan sumber penerimaan negara untuk menutupi seluruh pengeluaran, termasuk pembangunan. Kebijakan pemungutan pajak suatu negara harus mempunyai dasar hukum. Jika tidak, maka segala pungutan yang dilakukan oleh negara akan dianggap sebagai pungli (pungutan liar).

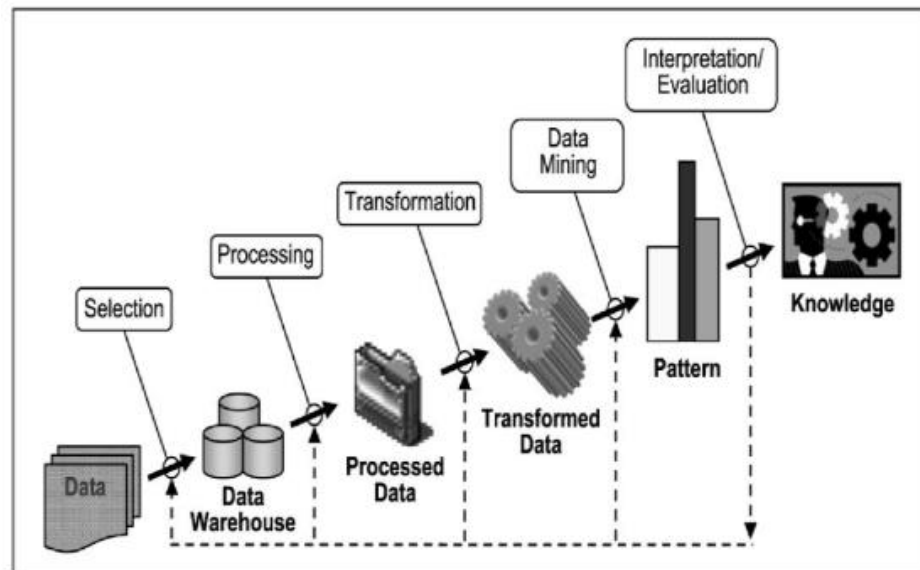
Salah satu manfaat pajak adalah mendanai pembangunan. Misalnya pajak bumi dan bangunan adalah pajak yang harus kita bayar karena kita tinggal di dalam negeri. Selain itu, pajak modal membiayai belanja modal, seperti biaya pegawai, barang, dan pembangunan fasilitas umum. Daerah juga menerima pajak dari pemerintah pusat untuk menunjang kegiatan pembangunan secara langsung dan tidak langsung (Mustaqiem, 2014).

Menurut Pasal 1 angka 2 Undang - Undang Nomor 28 Tahun 2007, Wajib Pajak adalah orang pribadi atau badan yang mempunyai hak dan kewajiban perpajakan sesuai dengan ketentuan peraturan perundang-undangan perpajakan. Undang-undang perpajakan mengatur hubungan hukum antara negara sebagai pemungut pajak dan masyarakat sebagai pembayar pajak. Hal ini menandakan terdapat dua badan hukum di bidang perpajakan masyarakat sebagai pembayar

pajak. Adapun fungsi pajak yakni : sebagai *Budgetair, regulierend*, Stabilitas, dan Redistribusi pendapatan. Apabila Wajib Pajak melakukan perubahan SPT berkala dan jumlah pajak yang dibayarkan bertambah, maka dikenakan pula denda administrasi sebesar 2 persen per bulan atas jumlah pajak yang kurang dibayar dihitung sejak batas waktu pembayaran sampai dengan saat pembayaran tanggal. Bagian-bagian suatu bulan dihitung sebagai satu bulan penuh (UU KUP, Pasal 8 ayat 2a). Meskipun telah dilakukan prosedur pemeriksaan, namun tidak dilakukan tindakan penyidikan terhadap adanya kebohongan yang dilakukan Wajib Pajak dalam pengertiannya (UU KUP, Pasal 38). Apabila kewajiban perpajakan terdiri dari kehendak bebas wajib pajak untuk mengungkapkan kepalsuan perbuatannya dan pelunasan kekurangan jumlah pajak yang sebenarnya, maka tidak dilakukan penyidikan mengenai kepalsuan perbuatan wajib pajak tersebut (Mustaqiem, 2014).

2.1.3. Data Mining

Data mining adalah istilah yang digunakan untuk menggambarkan penemuan pengetahuan dalam database serta, proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi berguna dan pengetahuan terkait dari database (Dicky Nofriansyah, 2015). Adapun istilah lain dari *data mining* yaitu *Knowledge discovery in database* (KDD) istilah ini sering digunakan untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*.



(sumber: Dicky Nofriansyah, 2015)

Gambar 2.1 Proses Knowledge discovery in database

a. *Data section*

Data section merupakan pemilihan atau seleksi data yang relevan dari sekumpulan data operasional, ini perlu dilakukan sebelum tahap penggalian informasi dalam KDD. Hasil data seleksi akan digunakan untuk proses data mining.

b. *Data Processing (Cleaning)*

Data pra-processing atau cleaning merupakan tahap pembersihan data yang tidak sempurna seperti kata-kata yang hilang, salah ketik dan juga atribut data yang tidak relevan. Dilakukan pembersihan data dikarenakan dapat mempengaruhi saat proses pemodelan dan akurasi data mining nantinya.

c. *Data Transformation*

Data Transformation merupakan proses pemilihan data yang diperlukan untuk proses data mining yang dipakai. Pada pemilihan data akan dirubah sesuai

bentuk data yang dapat digunakan oleh *tools* sehingga data digunakan pada proses data mining.

d. *Data Mining*

Data mining merupakan proses mencari atau pemilihan metode seperti karakterisasi, klasifikasi, *regresi*, *clustering*, *asosiasi*, dan lain lain atau pemilihan algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan.

e. *Evaluation*

Evaluation bisa disebut *interpretation* merupakan proses mengidentifikasi sejumlah pola pola yang dihasilkan oleh proses data mining. Pola ini pun dihasilkan dalam bentuk yang mudah dimengerti, serta pada proses *evaluation* dilakukan pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis.

2.1.4. Clustering

Clustering merupakan salah satu metode data mining untuk mengelompokkan data dengan karakteristik yang sama ke suatu kelompok dan data dengan karakteristik yang berbeda ke kelompok yang lain. Jenis pendekatan *Clustering* ada 2 yaitu metode clustering pendekatan partisi dan hirarki. *Clustering* dengan pendekatan hirarki, yang juga dikenal sebagai clustering dengan pendekatan partisi, mengelompokkan data dengan memilah-milah data yang telah dipelajari ke dalam *cluster* yang sudah ada. *Clustering* dengan pendekatan partisi, di sisi lain, mengelompokkan data dengan membuat suatu hirarki dalam bentuk dendogram, di mana data yang serupa akan ditempatkan pada hirarki yang berdekatan daripada

yang berjauhan (Ardilla dkk., 2021). Adapun jenis algoritma *clustering* seperti *Fuzzy C-Means*, *K-Means*, *K-Medoids*, dan lain-lain. Masing-masing algoritma memiliki keunggulan maupun kelemahan jika dibandingkan dengan algoritma lain, namun masing-masing metode data mining memiliki kesamaan prinsip yakni mengelompokkan data dengan mengukur jarak kesamaan dan mencari karakteristik yang sesuai dengan data dalam satu kelompok (Ena Tasia & Afdal, 2023).

Menurut (Irwansyah & Faisal, 2015) *clustering* sangat berguna dan bisa menemukan *group* atau kelompok yang tidak dikenal dalam data, karena objek yang didalam *cluster* memiliki kemiripan karakteristik antas satu sama lainnya dan berbeda dengan *cluster* yang lain. *Clustering* juga banyak digunakan pada *Business Intelligence*, pengenalan pola citra, *Web Search*, bidang Ilmu Biologi dan untuk keamanan (*Security*). *Clustering* juga dikenal sebagai data segmentasi karena *clustering* mempartisi banyak data set ke dalam banyak *group*.

2.1.5. Algoritma *K-Means*

K-Means merupakan algoritma dengan metode *unsupervised learning*. Algoritma ini mengelompokkan data berdasarkan kedekatannya terhadap titik tengah yang disebut pengelompokan dengan sistem partisi. *K-means clustering* yaitu teknik *clustering* non-hierarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*. Adapun aturan dalam pengelompokkan algoritma *K-Means* yaitu mengetahui jumlah pengelompokkan yang akan dibuat dan data hanya memiliki atribut numerik. Tujuan pengelompokkan *K-Means* untuk meminimalkan fungsi variasi antara data yang terdapat dalam suatu kelompok dan memaksimalkan variasi

dengan data kelompok lain. Berikut tahapan langkah-langkah *Clustering* Algoritma *K-Means*: (Ardilla dkk., 2021)

1. Menentukan jumlah klaster yang ingin dibentuk dan tetapkan pusat klaster k
2. Menentukan pusat *centroid* (klaster) secara acak
3. Setelah menentukan *centroid*, hitung jarak setiap data ke pusat *centroid* menggunakan persamaan *euclidean distance*.

Mengikuti persamaan 2.1 berikut:

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \dots\dots\dots (2.1)$$

Keterangan:

d_{ik} = jarak antara data ke- i ke *centroid cluster* ke- k

m = jumlah atribut

C_j = data *cluster* ke- j

C_k = data *cluster* ke- k

4. Mengelompokkan data berdasarkan yang paling dekat dengan *centroid*
5. Menghitung pusat *centroid* yang baru menggunakan persamaan 2.2 berikut:

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \dots\dots\dots (2.2)$$

Keterangan:

p = jumlah banyak *cluster*

x_{ij} = jumlah dari nilai jarak dalam masing-masing *cluster*

6. Melakukan perulangan langkah dua sampai langkah empat hingga tidak ada data yang berpindah ke *centroid* lain.

2.1.6. Algoritma *K-Medoids*

K-medoids merupakan algoritma pengelompokan yang lebih tahan terhadap *outlier*. Tujuan *K-medoids* adalah menemukan solusi pengelompokan yang meminimalkan fungsi yang telah ditentukan. Algoritma *K-medoids* lebih kuat terhadap noise dan *outlier* dalam data dan menggunakan titik data aktual sebagai prototipe. (Han dkk., 2012). Algoritma *K-Medoids* disebut *Partitioning Around*

Medoids (PAM). Sama seperti algoritma *k-means*, objek representatif awal dipilih secara acak. Serta mengevaluasi apakah penggantian objek representatif dengan objek non-representatif dapat meningkatkan kualitas pengelompokan. Semua penggantian yang mungkin dilakukan akan diuji. Proses iteratif penggantian objek representatif dengan objek lain akan terus berlanjut hingga tidak ada lagi penggantian yang dapat meningkatkan kualitas pengelompokan. Kualitas ini diukur dengan fungsi biaya dari rata-rata ketidakteraturan antara objek dan objek representatif dari klusternya (Aggarwal & Chandan, 2014).

Berikut langkah-langkah dalam metode *K-Medoids*: (Akbar dkk., 2023)

1. Tentukan jumlah kelompok k dan tentukan objek yang mewakili secara acak (medoid) sebanyak k .
2. Gunakan Persamaan (2.3) untuk menghitung jarak antara setiap objek dan objek representatif.

$$d(x_{ip}, o_{mp}) = \sqrt{\sum_{p=1}^n (x_{ip} - o_{mp})^2} \dots\dots\dots (2.3)$$

Keterangan:

$d(x_{ip}, o_{mp})$ = jarak antara objek ke- i x dan ke- m dalam variabel- p .

x_{ip} = objek ke- i x variabel ke- p

o_{mp} = variabel ke- p , objek medoid ke- m

3. Mengklasifikasikan setiap objek ke dalam kluster yang sesuai berdasarkan jarak terpendek dari medoid.
4. Temukan fungsi objektif, yang merupakan jumlah jarak terpendek dari medoid untuk setiap objek.
5. Pilih k objek yang tidak representatif, atau non-medoids.
6. Gunakan Persamaan (2.4) untuk menghitung jarak geometris antara setiap objek dan nonmedoids.

$$d(x_{ip}, o_{hp}) = \sqrt{\sum_{p=1}^n (x_{ip} - o_{hp})^2} \dots\dots\dots (2.4)$$

Keterangan:

$d(x_{ip}, o_{hp})$ = jarak antara objek ke-i x dari variabel-p dan objek ke-h yang bukan medoid dari variabel-p.

o_{hp} = objek non-medoid dengan nilai p ke-h

7. Tentukan apakah setiap objek termasuk dalam kelompok yang sesuai berdasarkan jarak terpendek dari non-medoids dan hitung fungsi objektif non-medoids.
8. Menghitung nilai total jarak baru dibagi dengan total jarak lama adalah cara untuk menghitung total simpangan (S). Jika S kurang dari 0, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid. Dengan rumus:

$$S = \text{total jarak baru} - \text{total jarak lama} \dots\dots\dots (2.5)$$

9. Teruskan melakukan langkah-langkah 5-8 sampai medoids tidak berubah.

2.1.7. Teori Slovin

Teori *Slovin* merupakan teknik penarikan sample *Purposive sampling* dengan metode kuantitatif, di mana sampel dipilih berdasarkan pertimbangan tertentu sample bila populasi jumlah data penelitian sangat besar mungkin ribuan orang atau lebih rumus pengambilan sampel tertentu digunakan untuk mengurangi jumlah sampel yang diambil atau wilayah populasi sehingga penelitian dapat dilakukan dengan lancar dan efisien (Fauzy, 2019). Pada prinsipnya, rumus pengambilan sampel digunakan untuk mempermudah teknis penelitian. Sebagai contoh, jika populasi penelitian sangat besar, seperti ribuan orang, atau wilayah populasi terlalu luas, rumus tertentu dimaksudkan untuk memperkecil jumlah sampel yang diambil atau mempersempit wilayah populasi sehingga penelitian

menjadi lancar dan efisien. Berikut perhitungan slovin dengan menggunakan persamaan 2.6 :

$$n = \frac{N}{1+N \cdot e^2} \dots\dots\dots (2.6)$$

Keterangan:

n = pengukuran sampel

N = pengukuran populasi

e = prosentase toleransi ketidaktekelitian yang disebabkan oleh kesalahan pengambilan sampel

2.1.8. Normalisasi

Normalisasi merupakan Mengubah nilai suatu variabel agar memiliki rentang atau skala yang konsisten, untuk mencegah variabel dengan rentang nilai yang luas, harus dinormalisasi atau distandarisasi. Normalisasi data memastikan bahwa semua atribut memiliki berat yang sama. Tujuan normalisasi data dalam data adalah untuk memastikan bahwa data dibentuk dalam rentang nilai yang sama memberikan bobot yang sama dalam semua atribut (Han dkk., 2012).

Normalisasi *Min-Max* merupakan metode normalisasi yang melakukan transformasi linier menggunakan nilai minimum dan maksimum. *Min-Max* menciptakan keseimbangan antara satu dan lainnya dalam rentang yang sama, dengan menggunakan persamaan (Puspitasari dkk., 2023). Normalisasi *Min-Max* dilakukan karena terdapat perbedaan interval dan ukuran angka yang sangat besar sehingga dapat mempersulit proses pengelompokan. Maka solusi yang digunakan untuk memperkecil ukuran angka antar variabel. Berikut untuk menghitung normalisasi min-max dengan menggunakan persamaan 2.7:

$$x \text{ normalized} = \frac{X-Xmin}{Xmax-Xmin} \dots\dots\dots (2.7)$$

Normalisasi *Z-score* merupakan metode normalisasi yang hasilnya diperoleh dari mean dan standar deviasi data. Metode ini kuat terhadap outlier dan

nilai yang lebih besar atau lebih kecil dari $\min A$ dan kurang dari $\max A$ atau $\min A$ (Saqila dkk., 2023). Berikut untuk menghitung normalisasi *z-score* dengan menggunakan persamaan 2.8:

$$Z = \frac{x - \tilde{x}}{\sigma} \dots\dots\dots (2.8)$$

Keterangan:

Z = nilai standar
 x = data mentah
 \tilde{x} = Nilai rata-rata
 σ = Simpang baku

2.1.9. Metode *Elbow*

Metode *Elbow* merupakan metode yang menghitung, menentukan jumlah *cluster* yang terbaik berdasarkan persentase perbandingan jumlah *cluster* yang membentuk busur pada suatu titik. Banyaknya klaster dikatakan benar jika nilai klaster pertama dan kedua membentuk sudut-sudut grafik, atau jika nilai klaster kedua mengalami penurunan paling besar (Maori & Evanita, 2023).

Untuk menghitung nilai metode elbow ini, menggunakan perhitungan *Sum Squared Error* (SSE) dengan Selisih total antara nilai yang dicapai dengan nilai sebenarnya SSE merupakan hasil penjumlahan seluruh jarak seluruh data ke pusat *cluster*. Nilai SSE yang rendah atau mendekati 0 menunjukkan bahwa hasil perhitungan memiliki nilai kesalahan acak yang minimal (Puspitasari dkk., 2023).

Berikut perhitungan *Sum Squared Error* (SEE) dapat dilihat pada persamaan 2.9

$$SSE = \sum_{i=1}^n (X_i - Y_i)^2 \dots\dots\dots (2.9)$$

2.1.10. Silhouette Coefficient

Silhouette Coefficient merupakan metode evaluasi kualitas baik atau buruk hasil dari *cluster* yang terbentuk. Nilai *siluet* digunakan sebagai alat untuk

mengukur kualitas penugasan kluster pada titik waktu tertentu. Skor yang positif menunjukkan kualitas penugasan yang baik, dengan skor yang lebih tinggi lebih diutamakan daripada skor yang lebih rendah. Nilai nol dianggap sebagai penugasan yang kurang kuat, karena observasi tersebut dapat ditempatkan di kluster terdekat berikutnya tanpa efek yang signifikan. Nilai siluet negatif dianggap tidak tepat; observasi akan lebih baik dilakukan di kluster terdekat lainnya (Larose & Larose, 2015). Berikut adalah langkah-langkah yang digunakan untuk menghitung *silhouette coefficient*:

1. Hitung jarak rata-rata antara objek, misalnya objek ke- i , dan semua objek lain yang berada dalam satu kelompok dengan Persamaan 2.10 berikut:

$$a(i) = \frac{1}{n_A - 1} \sum_{j \in A, j \neq i} d(i, j) \dots \dots \dots (2.10)$$

2. Menghitung jarak rata-rata dari objek ke- i ke semua objek pada *cluster* lain (misalnya, C adalah semua *cluster* yang dibentuk selain A), dapat dilakukan dengan menggunakan Persamaan (2.11) berikut:

$$b(i) = \min_{C \neq A} \left(\frac{1}{n_C} \sum_{j \in C} d(i, j) \right) \dots \dots \dots (2.11)$$

3. Nilai *siluet* dapat dihitung menggunakan Persamaan (2.12) berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i); b(i))} \dots \dots \dots (2.12)$$

4. Menghitung Koefisien Siluet (SC) dengan menggunakan Persamaan (2.13) berikut:

$$SC = \frac{1}{n} \sum_{i=1}^n s(i) \dots \dots \dots (2.13)$$

Nilai *Silhouette Coefficient* dapat menunjukkan seberapa dekat objek-objek dalam *cluster* mirip satu sama lain. Nilai *Silhouette Coefficient* yang lebih tinggi

akan menunjukkan hasil *clustering* yang lebih baik, sedangkan nilai *Silhouette Coefficient* yang lebih rendah akan menunjukkan hasil *clustering* yang lebih buruk.

Berikut tabel Kriteria *Silhouette Coefficient* ditunjukkan pada tabel 2.1

Tabel 2.1 Kriteria *Silhouette Coefficient*

<i>Silhouette Coefficient</i>	Kriteria
$0.7 < SC \leq 1$	Struktur yang kuat
$0.5 < SC \leq 0.7$	Struktur yang Baik
$0.25 < SC \leq 0.5$	Struktur Lemah
$0.25 \leq SC$	Tidak terstruktur

2.2. Kajian Penelitian

Beberapa penelitian terkait telah dilakukan sebelumnya, yang digunakan sebagai tolak ukur dan perbandingan dengan penelitian yang berjudul Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk *Clustering* Wilayah Dusun Wajib Pajak Pada Dokumen Letter C.

Adapun kajian penelitian yang bertema Letter C. Penelitian pertama yang dilakukan (Rahmi dkk., 2023) yang berjudul "Peningkatan Kualitas Pelaporan DHKP (Daftar Himpunan Ketetapan Pajak dan Pembayaran) Desa Kedungsolo Porong Sidoarjo Tahun 2022" menjelaskan Pajak Bumi dan Bangunan (PBB) ditetapkan berdasarkan tujuannya—tanah dan bangunan, bukan subjek pajak. Pembayaran diserahkan ke bank daerah oleh pemerintah daerah. Di Desa Kedungsolo, pemungutan pajak PBB masih dilakukan secara manual, dan pencatatan distribusi SPPT PBB masih menjadi masalah. Akibatnya, sangat sulit untuk menemukan wajib pajak. Untuk meningkatkan kualitas laporan DHKP, tujuannya adalah membuat sistem informasi dalam bentuk dashboard menggunakan PHP dan MySQL. Teknologi ini membantu pemerintah desa dan masyarakat karena memungkinkan pemantauan dan pencatatan PBB secara online serta

memungkinkan wajib pajak untuk melihat secara langsung status SPPT PBB mereka.

Penelitian kedua yang dilakukan (Santi dkk., 2023) yang berjudul ”*Engineering Design Business Process Modelling Letter C Land Data Archiving System with Software Requirement Specifications Approach*” menjelaskan Berdasarkan penelitian tersebut maka dibuatlah rancangan teknis sistem pengumpulan, pengambilan, dan pemetaan data huruf C di desa Pandanalum. Perancangan ini menggunakan model SDLC dengan pendekatan kebutuhan SRS. Hasil penelitian ini akan menjadi landasan bagi penelitian selanjutnya yang mempertimbangkan rekayasa bisnis dalam konteks merancang dan mengimplementasikan sistem nyata yang membantu perangkat desa memenuhi kebutuhannya.

Penelitian kedua yang dilakukan (Putri dkk., 2022) yang berjudul “Penerapan Data Mining Menggunakan Metode Clustering Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan” memiliki tujuan mengetahui tingkat kepatuhan Pajak Bumi dan Bangunan (PBB) di Desa Kendal, Kabupaten Cirebon. Pemerintah desa tidak dapat memastikan tingkat kepatuhan wajib pajak karena kurangnya informasi mengenai pembayaran langsung ke kas negara. Oleh karena itu, pada tahun 2021, data dari 1.159 wajib pajak di Desa Kendal dianalisis dengan menggunakan teknik data mining yang disebut K-Means clustering. Berdasarkan hasil analisis, *cluster 0* memiliki kepatuhan yang buruk, *cluster 1* memiliki kepatuhan yang sedang, *cluster 2* memiliki kepatuhan yang baik, dan *cluster 3* memiliki kepatuhan yang sangat tinggi dalam pembayaran PBB. Di

setiap klaster, penilaian PBB yang paling banyak dilakukan adalah dengan nilai rata-rata Rp 18.000 per klaster.

Penelitian ketiga (Umar & Irawan, 2023) yang berjudul "Aplikasi Pembayaran Pajak Bumi dan Bangunan Desa Tulus Rejo Kecamatan Pekalongan" menjelaskan Kemajuan teknologi, terutama yang melibatkan komputer, telah memudahkan berbagai pekerjaan. Tujuan dari penelitian ini adalah untuk mengembangkan sebuah aplikasi desktop untuk pembayaran pajak bumi dan bangunan di kantor balai desa Tulus Rejo. Aplikasi ini dimaksudkan untuk mempermudah proses pembayaran pajak, sehingga petugas dapat melayani wajib pajak dengan lebih efektif. Penulis menggunakan teknik Waterfall untuk memastikan pengembangan tahap demi tahap yang tidak ambigu, menggabungkan desain terstruktur dan Blackbox Testing untuk tujuan pengujian. Penelitian ini menghasilkan sebuah karya tulis skripsi yang berjudul "Aplikasi Pembayaran Pajak Bumi dan Bangunan di Kelurahan Tulus Rejo, Kabupaten Pekalongan

Penelitian keempat (Anggara & Nugroho, 2022) yang berjudul "Perancangan Sistem Informasi Pemantauan Pajak Bumi dan Bangunan Perdesaan dan Perkotaan (PBB-P2) Berbasis Web Menggunakan Metode Prototype di Kelurahan Parung Serab" menjelaskan Kelurahan Parung Serab, sebuah kecamatan di sebuah kota/kabupaten, mengalami kesulitan dalam melakukan pendataan dan pengelolaan pembayaran Pajak Bumi dan Bangunan Perdesaan dan Perkotaan (PBB-P2) secara manual karena proses input data yang memakan waktu lama dan pencatatan yang kurang efektif. Untuk mengatasi hal tersebut, dibuatlah sebuah sistem monitoring pajak berbasis web dengan menggunakan PHP dan MySQL.

Sistem ini memudahkan penginputan data pembayaran pajak, memudahkan pemantauan pembayaran PBB-P2, dan mengorganisir pendistribusian SPPT, sehingga mengurangi beban kerja administratif pihak kecamatan

Tabel 2.2 Kajian Penelitian Dokumen Ketetapan Pajak

Penulis	Judul	Hasil	Perbedaan
(Rahmi dkk., 2023)	Peningkatan Kualitas Pelaporan DHKP (Daftar Himpunan Ketetapan Pajak dan Pembayaran) Desa Kedungsolo Porong Sidoarjo Tahun 2022	Hasil penelitian ini memberikan manfaat bagi pemerintah desa dan masyarakat dengan memungkinkan pemantauan dan pencatatan PBB secara online, serta memungkinkan wajib pajak untuk melihat status SPPT PBB mereka secara individual	Pengembangan menggunakan PHP dan MySQL
(Santi dkk., 2023)	<i>Engineering Design Business Process Modelling Letter C Land Data Archiving System with Software Requirement Specifications Approach</i>	Hasil penelitian ini mencakup identifikasi analisis kebutuhan sistem yang fungsional dan non-fungsional serta perancangan sistem yang terdiri dari model proses bisnis dan use case diagram. Penemuan ini memberikan dasar untuk penelitian lanjutan.	Pengembangan sistem dilakukan dengan menggunakan <i>Software Development Life Cycle (SDLC)</i>
(Putri dkk., 2022)	Penerapan Data Mining Menggunakan Metode Clustering Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan	Hasil dari metode K-Means menunjukkan bahwa data Wajib Pajak Bumi dan Bangunan Desa Kendal pada tahun 2021 terbagi menjadi empat kelompok Blok. Pada kelompok 0 sebesar 77%, kelompok 1 sebesar 76%, kelompok 2 sebesar 76%, dan kelompok 3 sebesar 44,3%.	Variable SPPT kabupaten Cirebon Proses Cluster menggunakan 1 Algoritma
(Umar & Irawan, 2023)	Aplikasi Pembayaran Pajak Bumi dan Bangunan Desa Tulus RejoKecamatan Pekalongan	Hasil dari penelitian ini yaitu plikasi desktop yang dapat digunakan untuk pembayaran pajak bumi dan bangunan Desa Tulus Rejo	Aplikasi menggunakan PHP Metode Waterfall
(Anggara & Nugroho, 2022)	Perancangan Sistem Informasi Pemantauan Pajak Bumi dan Bangunan Perdesaan dan Perkotaan (PBB-P2) Berbasis Web Menggunakan Metode Prototype di	Hasil penelitian ini yakni mengembangkan sistem informasi pemantauan pajak bumi dan bangunan perdesaan dan perkotaan (PBB-P2) yang memungkinkan kelurahan Parung Serab untuk mencatat pembayaran pajak, melakukan pemantauan pembayaran PBB-P2, dan mencatat dokumen	Aplikasi Sistem Informasi menggunakan PHP dan MySQL Metode prototype dengan UML (<i>Unified Modeling Language</i>)

Kelurahan Parung Serab SPPT yang didistribusikan secara terorganisir.

Tabel 2.3 Kajian Penelitian Perbandingan Algoritma *Clustering*

Penulis	Judul	Hasil	Perbedaan
(Abbas dkk., 2020)	<i>K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir</i>	Hasil pengelompokan menggunakan Metode Transformasi Yeo-Johnson menunjukkan K-means K-medoids fungsi jarak SEV) dan Rank k-medoids (SEV fungsi jarak) dengan akurasi rata-rata 67,58%, 69,58%, dan 72,64%, masing-masing.	Variable Data rumah sakit pemerintah di kota Muzaffarabad, ibu kota Azad Kashmir Metode pengujian transformasi
(Novia & Sulastrri, 2022)	<i>Clustering Pop Songs Based On Spotify Data Using K-Means And K-Medoids Algorithm</i>	Hasil perhitungannya menunjukkan bahwa K means cluster 1 memiliki nilai 118, cluster 2 memiliki nilai 125, dan cluster 3 memiliki nilai 123. K medoids cluster 1 memiliki nilai 129, cluster 2 memiliki nilai 122, dan cluster 3 memiliki nilai 110.	Variable Data Spotify Metode CRISP-DM
(Karo Karo dkk., 2023)	<i>K-means and K-medoids Algorithm Comparison for Clustering Forest Fire Location in Indonesia</i>	Hasil dievaluasi dengan silhouette coefficient (SC), enam cluster adalah yang terbaik. K-medoids algoritma mengelompokkan data lebih baik dengan nilai 0,813 daripada K-means memiliki kurang dari 0,8.	Variable data lokasi hotspot Global Watch (GFW) Metode ini menggunakan sensor MODIS pada satelit Terra/Aqua dan satelit SNP VIIRS
(Lesmana & Gunawan, 2022)	Perbandingan Algoritma <i>K-Means</i> dan <i>K-Medoids</i> Dalam Penclusteran Data Penjualan PT. United Teknologi Integrasi	Hasil uji kualitas cluster metode Silhouette Index menghasilkan nilai maksimum 0.404 untuk algoritma K-Means dan 0.376 untuk algoritma K-Medoids.	Variable Penjualan PT. United Teknologi Integrasi
(Luchia dkk., 2022)	Perbandingan K-Means dan K-Medoids Pada Pengelompokan	Hasil terbaik dari algoritma. Menggunakan <i>Davies-Bouldin Index</i>	Variable Data Miskin

Penulis	Judul	Hasil	Perbedaan
	Data Miskin di Indonesia	melakukan pengujian ini dengan hasil terbaik, dengan K-Means s sebesar 0,041 dengan percobaan K=8 dan K-Medoids sebesar 0.052 dengan percobaan K=2.	Metode pengujian <i>Davies-Bouldin Index</i>
(Herviany dkk., 2021)	Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat	Hasil pengelompokan menggunakan <i>Davies-Bouldin Index</i> menghasilkan nilai. Nilai DBI terkecil untuk K-Means adalah 0,265 dengan k = 6, dan nilai DBI terkecil untuk K-Medoids adalah 0,342.	Variable data tanah longsor pada Provinsi Jawa Barat Metode pengujian DBI
(Fitriyadi & Kurniawati, 2021)	Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional	Hasil pengelompokan menggunakan metode <i>Confusion Matrix</i> dengan parameter precision, recall, dan akurasi. Menurut penelitian, algoritma K-Means memiliki akurasi 56%, presisi 25%, dan recall 60%, sedangkan algoritma K-Medoids memiliki akurasi 14%, presisi 25%, dan recall 25%.	Variable data kinerja karyawan Metode <i>Confusion Matrix</i> dengan parameter <i>accuracy</i> , <i>recall</i> dan <i>precision</i>
(Puspitasari dkk., 2023)	Perbandingan Algoritma K-Means dan Algoritma K-Medoids Pada Kasus Covid-19 di Indonesia	Hasil pengelompokan ini dibuktikan dengan nilai SC pada kasus positif, sembuh, dan meninggal sebesar 0.838; 0.838; dan 0.925, sedangkan nilai DBI sebesar 0.305 untuk kasus positif, 0.295 untuk kasus sembuh, dan 1.569 untuk kasus meninggal. Selanjutnya, nilai SSE adalah 3.381, sembuh 3.251, dan meninggal 4.049.	Variable Kasus Covid 19 di Indonesia Metode Sum Squared Error (SSE) dan Davies Index Bouldin (DBI).
(Farissa dkk., 2021)	Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient	Hasil clustering K-Means lebih baik, karena Silhouette Coefficient lebih tinggi sebesar 0,627 sedangkan K-Medoids sebesar 0,536.	Variable data Obat Metode <i>Cross-Industry Standard Process for Data mining (CRISP-DM)</i>
(Fira dkk., 2021)	Komparasi Algoritma K-Means dan K-Medoids Untuk Pengelompokan Penyebaran Covid-19 di Indonesia	Hasil perbandingan menghasilkan nilai koefisien silhouette sebesar 0,207 dengan metode K-Means dan nilai koefisien silhouette sebesar 0,347 dengan metode K-Medoids.	Variable data penyakit Covid-19 di indonesia tahun Desember 2019 - Februari 2021

Penulis	Judul	Hasil	Perbedaan
(Supriyadi dkk., 2021)	Perbandingan Algoritma K-Means dengan K-Medoids pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas	Hasil validitas cluster yang terbentuk yang diuji menggunakan <i>Davies-Bouldin Index</i> (DBI) dengan Rapid Miner; validitas untuk clustering K-Means adalah 0,67 dan untuk K-Medoids adalah 1,78, masing-masing.	Variabel data armada kendaraan Metode pengujian <i>Davies-Bouldin Index</i>
(Wahyudi & Pujiastuti, 2022)	Komparasi K-Means Clustering dan K-Medoids dalam Mengelompokkan Produksi Susu Segar di Indonesia	Hasil Validitas Evaluasi nilai DBI untuk algoritma K-Means Clustering adalah 0,094 dan K-Medoids Clustering adalah 0,072	Variabel data produksi susu segar Metode pengujian <i>Davies-Bouldin Index</i>
(Hidayat & Fitriana, 2022)	Implementasi K-Means Dan K-Medoids Dalam Pengelompokan Wilayah Potensial Produksi Daging Ayam	Hasil dari pengujian DBI pada algoritma K-Mean = 0,273 dan DBI pada algoritma K-Medoids = 0,598, jumlah k yang ideal adalah lima kelompok. Hasil clustering menunjukkan bahwa daerah produksi daging ayam di provinsi Jawa Barat terbagi menjadi lima kelompok.	Variabel data produksi daging ayam Metode pengujian <i>Davies-Bouldin Index</i>
(Prasetyaningrum & Susanti, 2023)	Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan	Hasil dari perngujian menggunakan DBI yaitu Algoritma K-Means mendapatkan nilai DBI (Davies Bouldin Index) yang lebih rendah sebesar 0,296, sedangkan algoritma K-Medoids mendapatkan nilai DBI (Davies Bouldin Index) sebesar 0,507.	Variable data produksi buah-buahan Metode pengujian <i>Davies-Bouldin Index</i>
(Zahrotun dkk., 2023)	<i>Comparison of K-Medoids Method and Analytical Clustering on Students' Data Grouping</i>	Hasil menunjukkan nilai linier dapat siluet linier antara metode AHC dan K-Medoids, pendekatan AHC memiliki nilai siluet yang lebih baik (jurusan R: 0.35, jurusan S: 0.65 jumlah klaster 2, jurusan T: 0.67 jumlah klaster 2, dan program studi U: 0.52). Namun, nilai siluet pendekatan AHC lebih tinggi (jurusan R: 0.88, jurusan S: 0.87, jurusan T: 0.88, dan program studi U: 0.88).	Variable data mahasiswa di universitas X Algoritma menggunakan AHC
(Ena Tasia & Afdal, 2023)	Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering	Hasil Dalam percobaan yang dilakukan dengan alat RapidMiner ditunjukkan bahwa algoritma K-Means	Variable data daerah rawan banjir

Penulis	Judul	Hasil	Perbedaan
	Daerah Rawan Banjir Di Kabupaten Rokan Hilir	memiliki kluster optimal dengan $k = 3$, dengan nilai validitas Davies-Bouldin Index (DBI) sebesar 0.218%, sedangkan algoritma K-Medoids memiliki kluster optimal dengan $k = 4$, dengan nilai validitas Davies-Bouldin Index (DBI) 0.525%.	Metode pengujian <i>Davies-Bouldin Index</i>
(Nowak, 2023)	<i>Comparative analysis of selected algorithms for qualitative data clustering</i>	Hasil dari bahwa algoritma k-modes memberikan hasil yang lebih baik dalam hal kualitas pengelompokan dan dengan demikian lebih efektif dalam mendistribusikan observasi ke dalam kelas-kelas.	Variable data <i>qualitative</i> Metode algoritma k-modes
(Nasari dkk., 2023)	Optimasi Metode K-Means dan K-Medoids Berdasarkan Jumlah Cluster dan Nilai DBI Dalam Pengelompokkan Produksi Kelapa Sawit Di Provinsi Riau	Hasil penelitian menunjukkan bahwa cluster 2 adalah yang terbaik dengan nilai DBI tertinggi untuk metode K-medoid, yang mencapai 159796492242,667 dan metode K-Means, yang mencapai 82338884292,014. Jumlah cluster yang dihasilkan menunjukkan bahwa metode K-medoid adalah yang terbaik karena menghasilkan 7 kabupaten di kelompok jumlah produksi Tinggi dan 5 kabupaten di kelompok jumlah produksi Rendah.	Variabel data BPS provinsi riau, data produksi perkebunan tahun 2018-2019 Metode pengujian <i>Davies-Bouldin Index</i>
(Fathia Palembang dkk., 2022)	Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokkan Tingkat Kebahagiaan Di Provinsi Indonesia	Hasil penelitian menunjukkan bahwa Algoritma k-Medoids mengungguli algoritma k-Means dan menghasilkan dua kluster—(kluster 0) dan kluster 1)—berdasarkan indeks kepuasan, emosi, dan makna hidup.	Variable data data indeks kebahagiaan tahun 2021 indonesia Metode pengujian <i>Davies-Bouldin Index</i>
(Qomariyah & Siregar, 2022)	<i>Comparative Study of K-Means Clustering Algorithm and K-Medoids Clustering in Student Data Clustering</i>	Hasil dari proses K-Means kluster 5 dengan nilai DBI sebesar 0,781. Pada proses K-Medoids mendapatkan jumlah kluster terbaik yaitu 3 dengan nilai DBI sebesar 0,929. Berdasarkan nilai uji validasi DBI bahwa algoritma K-Means lebih optimal dari K-Medoids.	Variable data siswa Metode pengujian <i>Davies-Bouldin Index</i>

Penulis	Judul	Hasil	Perbedaan
		Sehingga klaster mahasiswa dengan rata-rata IPK tertinggi sebesar 3,325 sejumlah 401 mahasiswa.	

Penelitian ini dilakukan sebagai bahan perbandingan dari penelitian sebelumnya. Adapun hasil penelitian yang dijadikan perbandingan tetap terkait dengan topik penelitian analisis Data mining, metode K-Means dan K-Medoids .

Penelitian pertama oleh (Abbas dkk., 2020) yang berjudul "*K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir*" menjelaskan Dalam industri medis, menilai statistik perawatan kesehatan untuk wanita hamil di negara berpenghasilan rendah sangat penting karena tingginya angka kematian akibat operasi caesar yang disebabkan oleh infrastruktur yang tidak memadai, keyakinan agama yang salah, dan kurangnya pengetahuan. Dilakukan secara lokal di Muzaffarabad, Kashmir, proyek ini bertujuan untuk mendapatkan informasi dari ibu hamil untuk membantu mereka dalam membuat keputusan medis. Selain teknik transformasi data seperti skala, rentang, dan Yeo-Johnson, algoritma K-means dan K-medoids dipelajari. Pendekatan transformasi Yeo-Johnson memberikan hasil terbaik untuk algoritma K-means, K-medoids, dan Rank K-medoids, masing-masing dengan akurasi rata-rata 67,58%, 69,58%, dan 72,64%, masing-masing.

Penelitian kedua oleh (Novia & Sulastri, 2022) yang berjudul "*Clustering Pop Songs Based On Spotify Data Using K-Means And K- Medoids Algorithm*" menjelaskan hasil pengelompokan algoritma K-Means dan K-Medoids dengan alat Rstudio pada kualitas musik Spotify seperti Danceability, Energy,

Acousticness, Instrumentalness, Liveness, Loudness, Speechiness, Valence, dan Tempo. Jumlah cluster yang sesuai telah ditentukan, serta hasil cluster tinggi, sedang, dan rendah. Nilai tempo tertinggi untuk algoritma K-Means adalah 118, 125, dan 123 untuk algoritma K-Medoids untuk cluster 1, 2, dan 3.

Penelitian ketiga oleh (Karo Karo dkk., 2023) yang berjudul "*K-means and K-medoids Algorithm Comparison for Clustering Forest Fire Location in Indonesia*" menjelaskan Deforestasi di Indonesia disebabkan oleh kebakaran hutan, ini membahayakan keanekaragaman hayati. Algoritma K-Means dan K-Medoids digunakan untuk mengelompokkan titik api untuk menghindari kebakaran ini. Sembilan faktor penting diperiksa, dengan Acq_time yang paling relevan, dari data Global Forest Watch (GFW). Kedua metode berkinerja baik dalam pengelompokan, dengan K-Medoids mengalahkan K-Means dalam pengelompokan data, dengan enam kelompok yang ideal.

Penelitian keempat oleh (Lesmana & Gunawan, 2022) yang berjudul "*Perbandingan Algoritma K-Means dan K-Medoids Dalam Penclusteran Data Penjualan PT. United Teknologi Integrasi*" menjelaskan Pencatatan transaksi penjualan sangat penting untuk melacak penjualan produk dan meningkatkan keuntungan bisnis. Data penjualan PT United Teknologi Integrasi dikumpulkan menggunakan algoritma K-Means dan K-Medoids melalui pendekatan data mining. Untuk menemukan kelompok ideal, data dari tahun 2018–2020 dipelajari. Ini menghasilkan tiga kelompok: barang terlaris, cukup laris, dan kurang populer. Temuan ini memberikan informasi tentang promosi yang dimaksudkan untuk meningkatkan penjualan. Dengan nilai indeks maksimum 0,404 untuk K-Means

dan 0,376 untuk K-Medoids, uji kualitas kluster dengan metode Silhouette Index menunjukkan bahwa K-Means lebih baik dalam mengelompokkan data penjualan.

Penelitian kelima oleh (Luchia dkk., 2022) yang berjudul "Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia" menjelaskan Di Indonesia, kemiskinan adalah masalah yang signifikan dan semakin mengkhawatirkan. Pemerintah masih tidak cukup berkomitmen untuk mengakhiri kemiskinan, yang berdampak pada bidang ekonomi, sosial, dan politik. Bantuan langsung tunai (BLT) dan inisiatif lainnya adalah bagian dari upaya untuk mengurangi kemiskinan. Tujuan dari penelitian ini adalah untuk mengumpulkan data statistik tentang penduduk miskin untuk memberikan wawasan kepada pemerintah tentang daerah-daerah dengan tingkat kemiskinan tertinggi. Data Badan Pusat Statistik (BPS) menunjukkan bahwa Indonesia memiliki 26,50 juta jiwa, atau sekitar 9,7% dari populasi dunia yang hidup dalam kemiskinan dari tahun 2014 hingga 2021. Seperti yang ditunjukkan oleh skor DBI sebesar 0,041 untuk $K = 8$, penelitian ini menunjukkan bahwa algoritma K-Means lebih baik dalam mengelompokkan data yang lemah daripada K-Medoids.

Penelitian keenam oleh (Herviany dkk., 2021) yang berjudul "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat" menjelaskan Indonesia sering mengalami bencana alam seperti banjir dan tanah longsor, yang sering diperparah oleh aktivitas manusia. Para peneliti menemukan 609 kasus tanah longsor di Jawa Barat pada tahun 2019 dengan menggunakan data Open Data. Badan Penanggulangan Bencana Daerah (BPBD) mengalami kesulitan untuk

mengoptimalkan layanan karena sumber daya yang tidak mencukupi. Pemetaan risiko bencana sangat penting untuk manajemen bencana yang efektif. Dengan enam klaster menjadi jumlah yang paling optimal, metode K-Means mengungguli metode K-Medoids dalam evaluasi data tanah longsor di Jawa Barat. Cluster 2 memiliki wilayah terluas, dan Cluster 5 memiliki jumlah kejadian tertinggi (106 di empat wilayah).

Penelitian ketujuh oleh (Fitriyadi & Kurniawati, 2021) yang berjudul “Analisis Algoritma *K-Means* dan *K-Medoids* Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional” menjelaskan Perumahan Nasional menilai kinerja seluruh karyawan setiap enam bulan sekali. Studi ini menguji algoritma *K-Means* dan *K-Medoids* ketika mereka membagi karyawan ke dalam empat kategori kinerja: sangat baik, baik, cukup, dan buruk. Perangkat lunak Rapidminer digunakan untuk mengelompokkan, dan pendekatan *Confusion Matrix*, yang menilai presisi, akurasi, dan recall, digunakan untuk mengukur kinerja. Hasilnya menunjukkan bahwa *K-Means* lebih baik daripada *K-Medoids* dalam hal akurasi (lima puluh enam persen) dan recall (enam puluh persen). Akibatnya, *K-Means* lebih baik daripada algoritma pengelompokan lainnya untuk menilai kinerja karyawan.

Penelitian kedelapan oleh (Puspitasari dkk., 2023) yang berjudul ” Perbandingan Algoritma K-Means dan Algoritma K-Medoids Pada Kasus Covid-19 di Indonesia” menjelaskan penelitian menggunakan metode pengelompokan K-Means dan K-Medoids untuk membagi contoh positif ke dalam tiga kategori: rendah, sedang, dan tinggi, penelitian ini menyelidiki data COVID-19 dari semua

provinsi di Indonesia. Hasil penelitian menunjukkan bahwa metode penghitungan jarak Manhattan Distance adalah yang terbaik; hasilnya menunjukkan bahwa K-Means mengalahkan K-Medoids dengan nilai Silhouette Coefficient dan Index Davies Bouldin yang lebih tinggi, menunjukkan kemampuan pengelompokan yang lebih baik.

Penelitian kesembilan oleh (Farissa dkk., 2021) yang berjudul ” Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient” menjelaskan Perencanaan kebutuhan obat yang efektif dan efisien diperlukan untuk mengatasi masalah distribusi obat yang tidak merata di tempat pelayanan kesehatan. Data mining dapat digunakan untuk mengontrol konsumsi obat untuk mencegah penimbunan atau kehabisan obat. Di Puskesmas Karangasambung, algoritma K-Means dan K-Medoids digunakan untuk mengelompokkan data obat. Hasil menunjukkan bahwa K-Means memiliki Silhouette Coefficient sebesar 0.627, sedangkan K-Medoids memiliki nilai 0.536. Pengelompokan ini memungkinkan fasilitas kesehatan untuk mendistribusikan obat secara merata, mencegah kelebihan atau kekurangan.

Penelitian kesepuluh oleh (Fira dkk., 2021) yang berjudul ” Komparasi Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Penyebaran Covid-19 di Indonesia” menjelaskan Virus COVID-19, yang berhubungan dengan SARS dan MERS, menyebabkan batuk, demam, kelelahan, dan kehilangan nafsu makan. Dalam penelitian ini, data mining digunakan untuk mengelompokkan 34 provinsi di Indonesia berdasarkan tingkat COVID-19. Algoritma K-Means dan K-Medoids digunakan untuk melakukan ini, menghasilkan dua cluster (2 tinggi, 32 rendah),

sementara K-Medoids menghasilkan dua cluster (4 tinggi, 30 rendah), dengan nilai Silhouette Coefficient masing-masing 0,207 dan 0,347.

Penelitian kesebelas oleh (Supriyadi dkk., 2021) yang berjudul ” Perbandingan Algoritma K-Means dengan K-Medoids pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas” menjelaskan Produksi transportasi bergantung pada manajemen armada yang efektif. Metode evaluasi manual dapat memberikan hasil yang rumit dan keliru, sehingga diperlukan metode pemrosesan data yang lebih cepat dan tepat, seperti pengelompokan data mining. Studi ini menyelidiki algoritma K-Means dan K-Medoids yang digunakan untuk mengelompokkan mobil berdasarkan kinerja dan produktivitas. K-Means menerima nilai validitas 0,67 dari Davies Bouldin Index, sedangkan K-Medoids menerima nilai validitas 1,78, sehingga K-Means dipilih untuk aplikasi pengelompokan armada berbasis web. Dengan menggunakan alat Rapidminer dan perhitungan manual, pengujian pada aplikasi web menunjukkan tingkat kesesuaian 97%.

Penelitian kedua belas oleh (Wahyudi & Pujiastuti, 2022) yang berjudul ” *Comparison of K-Means Clustering and K-Medoids in Clustering Fresh Milk Production in Indonesia*” penelitian ini membandingkan algoritma K-Means dan K-Medoids untuk mengkategorikan data produksi susu segar di Indonesia. Berdasarkan data dari tahun 2018–2020, K-Medoids unggul dengan nilai DBI yang lebih rendah (0,072). Upaya pemerintah Indonesia untuk meningkatkan produksi susu segar ditunjukkan oleh kelompokan produksi susu segar yang ideal.

Penelitian ketiga belas oleh (Hidayat & Fitriana, 2022) yang berjudul “Implementasi K-Means Dan K-Medoids Dalam Pengelompokan Wilayah Potensial Produksi Daging Ayam” menjelaskan Peternakan memainkan peran penting dalam pasokan makanan Indonesia, terutama dalam menjamin ketersediaan makanan hewani. Studi ini menggunakan algoritma K-Means dan K-Medoids untuk mengkategorikan lokasi penghasil daging ayam di seluruh provinsi Jawa Barat. Algoritma yang optimal dipilih dengan menggunakan Davis Bouldin Index (DBI). K-Means, dengan 5 cluster, secara akurat mengkategorikan potensi lokasi produksi daging ayam di Jawa Barat, dengan nilai DBI sebesar 0,273. Hasil pengelompokan ini dapat membantu mendorong rencana untuk meningkatkan produksi pakan hewani di daerah tersebut, meningkatkan potensi peternakan ayam, dan meningkatkan distribusi pakan ternak.

Penelitian keempat belas oleh (Prasetyaningrum & Susanti, 2023) yang berjudul “Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan” menjelaskan Dari 2019 hingga 2020, produksi buah di Kotawaringin Timur turun, menyebabkan pasokan terbatas dan kenaikan harga. Untuk mengatasi hal ini, algoritma K-Means dan K-Medoids digunakan untuk mengelompokkan data produksi buah. Tiga klaster dibuat untuk membantu Dinas Ketahanan Pangan dan Pertanian meningkatkan produksi pertanian. Dengan nilai DBI 0,507, algoritma K-Means mengalahkan K-Medoids dalam pengujian yang dilakukan dari tahun 2019 hingga 2022 dengan 29 titik data. Oleh karena itu, K-Means adalah pendekatan yang paling cocok untuk mengatur produksi buah di Kabupaten Kotawaringin Timur.

Penelitian kelima belas oleh (Zahrotun dkk., 2023) yang berjudul ”*Comparison of K-Medoids Method and Analytical Hierarchy Clustering on Students' Data Grouping*” menjelaskan penelitian ini membandingkan Analytic Hierarchy Clustering (AHC) dan K-Medoids untuk mengelompokkan data siswa Universitas X berdasarkan berbagai fitur. Nilai silhouette AHC lebih tinggi daripada K-Medoids (R: 0.35, S: 0.65 dengan dua cluster, T: 0.67 dengan dua cluster, dan U: 0.52) dengan teknik silhouette coefficient dengan 2, 3, dan 4 cluster. Keputusan antara AHC dan K-Medoids didasarkan pada distribusi data daripada jumlah data atau pengelompokan.

Penelitian ke enam belas oleh (Ena Tasia & Afdal, 2023) yang berjudul ” Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir” menjelaskan Sejak tahun 2005, banjir telah menjadi masalah rutin di Kabupaten Rokan Hilir. Banjir terburuk terjadi antara tahun 2011 dan 2014, merusak banyak tempat dan orang. Dinas sosial sangat penting dalam memberikan bantuan cepat kepada korban banjir. Untuk memberikan bantuan yang efektif, lokasi yang rawan banjir harus dikelompokkan. Studi ini mengelompokkan lokasi di Rokan Hilir yang rawan banjir dengan menggunakan algoritma K-Means dan K-Medoids. Hasil penelitian menunjukkan bahwa metode K-Means adalah yang terbaik, dengan konfigurasi cluster dengan $k=3$ dengan nilai validitas 0,218% Davies-Bouldin Index (DBI). Sebaliknya, metode K-Medoids mencapai konfigurasi cluster terbaik dengan $k=4$, dengan nilai validitas 0,525%.

Penelitian ke tujuh belas oleh (Nowak, 2023) yang berjudul ”*Comparative analysis of selected algorithms for qualitative data clustering*” menjelaskan Tujuan

pengelompokan adalah untuk membandingkan proses pengelompokan, kualitas pengelompokan yang dihasilkan, dan hasil pengelompokan antara yang dapat dituliskan dalam bentuk angka. Tujuan utama dari penelitian ini adalah untuk membandingkan proses clustering, kualitas cluster yang dihasilkan, dan hasil clustering antara yang dapat dituliskan dalam bentuk angka. K-means dan k-modes adalah algoritma yang digunakan untuk mengelompokkan pengamatan ke dalam kelas, dan keduanya digunakan pada tiga dataset. K-means dan k-modes memberikan kualitas pengelompokan yang lebih baik, sehingga lebih efektif dalam mendistribusikan pengamatan ke dalam kelas. tiga ukuran penentuan jarak yang berbeda digunakan. Ukuran Euklidis, ukuran perkotaan, dan ukuran Chebyshev adalah ukuran yang digunakan dalam eksperimen desain untuk mengelompokkan set numerik dan set data kualitatif menggunakan teknik pengkodean. Untuk struktur dari ketiga dataset yang dipilih, dapat disimpulkan bahwa algoritma mode k memberikan hasil yang lebih baik dalam hal kualitas pengelompokan dan dengan demikian lebih efektif dalam mendistribusikan observasi ke dalam kelas-kelas.

Penelitian kedelapan belas oleh (Nasari dkk., 2023) yang berjudul ” Optimasi Metode K-Means dan K-Medoids Berdasarkan Jumlah Cluster dan Nilai DBI Dalam Pengelompokkan Produksi Kelapa Sawit Di Provinsi Riau” Provinsi Riau, Indonesia, terkenal dengan 2,8 juta hektar perkebunan kelapa sawit yang menghasilkan 8,8 juta ton kelapa sawit setiap tahun. Teknik pengelompokan seperti K-means dan K-medoids digunakan untuk mengelola area ini dengan baik. K-medoids melampaui K-means dalam hasil, menurut pengujian dengan Rapid Miner

Tools. Hasilnya, lima kabupaten dianggap sebagai daerah produksi rendah, dan tujuh dianggap sebagai daerah produksi tinggi.

Penelitian ke-sembilan belas oleh (Fathia Palembang dkk., 2022) yang berjudul "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokan Tingkat Kebahagiaan Provinsi Di Indonesia" menjelaskan penelitian ini Fokus pengukuran keberhasilan pembangunan suatu wilayah berubah dari pengukuran pertumbuhan ekonomi dan kemiskinan ke pengukuran kebahagiaan masyarakat dan kesejahteraan. Dalam penelitian Indonesia baru-baru ini, algoritme k-Means dan k-Medoids dari RapidMiner Studio digunakan untuk mengelompokkan provinsi berdasarkan metrik kebahagiaan. Algoritma k-Medoids mengungguli algoritma k-Means dan menghasilkan dua klaster—klaster 0) dan klaster 1)—berdasarkan indeks kepuasan, emosi, dan makna hidup.

Penelitian ke-dua puluh oleh (Qomariyah & Siregar, 2022) yang berjudul "*Comparative Study of K-Means Clustering Algorithm and K-Medoids Clustering in Student Data Clustering*" menjelaskan Dalam penelitian ini, algoritma clustering K-Means dan K-Medoids diterapkan dan dianalisis pada data akademik mahasiswa UIN Sunan Kalijaga angkatan 2013–2015. Hasil penelitian menunjukkan bahwa metode K-Means dengan lima kelompok memiliki nilai Davies Bouldin Index (DBI) sebesar 0.781, sedangkan K-Medoids dengan tiga kelompok memiliki nilai DBI sebesar 0.929. Berdasarkan DBI, metode K-Means lebih optimal daripada K-Medoids, dengan klaster siswa tertinggi memiliki IPK rata-rata 3,325 dan terdiri dari 401 siswa.

Berdasarkan hasil dari beberapa penelitian yang dijadikan kajian Penelitian. Peneliti mengambil dua referensi jurnal atau artikel sebagai dasar dalam melakukan perbandingan *K-Means* dan *K-Medoids Clustering* Wilayah Dusun Berdasarkan Pajak Terhutang dengan menggunakan pengujian *Silhouette Coefficient* yakni:

Penelitian pertama oleh (Lesmana & Gunawan, 2022) yang berjudul "Perbandingan Algoritma *K-Means* dan *K-Medoids* Dalam Penclusteran Data Penjualan PT. United Teknologi Integrasi" menjelaskan Pencatatan transaksi penjualan sangat penting untuk melacak penjualan produk dan meningkatkan keuntungan bisnis. Data penjualan PT United Teknologi Integrasi dikumpulkan menggunakan algoritma *K-Means* dan *K-Medoids* melalui pendekatan data mining. Untuk menemukan kelompok ideal, data dari tahun 2018–2020 dipelajari. Ini menghasilkan tiga kelompok: barang terlaris, cukup laris, dan kurang populer. Temuan ini memberikan informasi tentang promosi yang dimaksudkan untuk meningkatkan penjualan. Dengan nilai indeks maksimum 0,404 untuk *K-Means* dan 0,376 untuk *K-Medoids*, uji kualitas klaster dengan metode *Silhouette Index* menunjukkan bahwa *K-Means* lebih baik dalam mengelompokkan data penjualan

Selanjutnya, Penelitian kedua yang dilakukan oleh (Fira dkk., 2021) yang berjudul "Komparasi Algoritma *K-Means* dan *K-Medoids* Untuk Pengelompokkan Penyebaran Covid-19 di Indonesia" menjelaskan Virus COVID-19, yang berhubungan dengan SARS dan MERS, menyebabkan batuk, demam, kelelahan, dan kehilangan nafsu makan. Dalam penelitian ini, data mining digunakan untuk mengelompokkan 34 provinsi di Indonesia berdasarkan tingkat COVID-19. Algoritma *K-Means* dan *K-Medoids* digunakan untuk melakukan ini, menghasilkan

dua cluster (2 tinggi, 32 rendah), sementara K-Medoids menghasilkan dua cluster (4 tinggi, 30 rendah), dengan nilai Silhouette Coefficient masing-masing 0,207 dan 0,347