BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian ini dibuat berdasarkan penelitian yang sudah dilakukan atau penelitian terdahulu yaitu penelitian berjudul "Penerapan Metode *K-Means* Untuk Menganalisis Minat Nasabah Asuransi" Berdasarkan penelitian sebelumnya bahwa Berdasarkan perhitungan data yang sudah dilakukan dengan menggunakan metode *K-Means* diperoleh 3 *cluster* dari keseluruhan data, yaitu *cluster* 1 untuk asuransi kebakaran mempunyai jumlah anggota sebanyak 30 orang, *cluster* 2 untuk asuransi kecelakaan mempunyai jumlah anggota sebanyak 24 orang, *cluster* 3 untuk asuransi kesehatan mempunyai jumlah anggota sebanyak 1 orang. Pada hasil akhir terlihat jelas semua data memiliki kedekatan yang sama antara satu dengan lainnya sehingga terbentuk satu pengelompokan berdasarkan jarak kedekatan dengan nilai data. Dengan demikian maka pembentukan *cluster* yang optimal dalam klasterisasi produk asuransi adalah menggunakan metode *K-Means*.[2]

Pada penelitian lainnya yang berjudul "Penerapan Data Mining Clustering Dengan Menggunakan Algoritma K-Means Pada Data Nasabah Kredit Bermasalah PT. BPR Milala" Berdasarkan hasil analisa dari permasalahan yang terjadi dengan kasus yang dibahas tentang mengelompokkan data nasabah dengan menerapkan metode terhadap k-means sistem yang dirancang dan dibangun maka dapat ditarik kesimpulan bahwa Dengan menganalisa pengelompokkan dalam kredit bermasalah dilakukan dengan riset ke tempat perusahaan dan wawancara

pihak pengolahan data nasabah yang dikelompokan dalam kredit bermasalah. Dengan menerapkan data mining dengan melakukan normalisasi data nasabah dan melakukan proses algoritama untnuk mendapatkan hasil keputusan dengan aplikasi dengan cepat ataupun akurat.[3]

Pada penelitian selanjutnya yang berjudul "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food" Mengetahui seberapa tertarik konsumen untuk membeli suatu produk dapat dilakukan dengan menghitung jumlah transaksi penjualan yang dilakukan, yang merupakan salah satu informasi yang dapat dikumpulkan. Sehingga semakin banyaknya kegiatan transaksi oleh konsumen terdapat data yang sangat besar dan banyak. Dari hasil penelitian *K-Means Clustering* untuk pengelompokan minat konsumen pada produk frozen food yang telah diuraikan disimpulkan bahwa penerapan algoritma *K-Means Clustering* pada data penjualan produk Frozen Food, menghasilkan sebuah informasi mengenai data pengelompokan minat konsumen tertinggi dan terendah. Dari 45 produk yang diteliti terdapat 3 produk yang merupakan anggota *Cluster* 1 atau dapat diartikan sebagai produk *frozen food* yang memiliki minat konsumen rendah dan 42 produk yang masuk ke dalam *Cluster* 2 yang dapat diartikan sebagai produk *frozen food* yang memiliki minat konsumen tinggi [4].

Pada penelitian selanjutnya tentang "implementasi metode K-Means Untuk Memprediksi Status Kredit Macet" Implementasi dari hasil metode K-Means memberikan hasil klasifikasi pengelompokan data yang efektif. Dari hasil cluster model penggunaan Rapidminer dengan hasil perhitungan manual memiliki hasil yang tidak jauh berbeda. Dengan hasil akhir penelitian menggunakan metode

k-means dapat menghasilkan data pengelompokan menjadi 3 kriteria, yaitu (C0) sebanyak 69 data dengan Nasabah Lancar, (C1) sebanyak 3 data dengan nasabah sangat lancar, dan (C2) sebanyak 52 data dengan nasabah macet. Dengan adanya penerapan metode k-means dengan menggunakan tools rapidminer dapat membantu PT. Esta Dana Ventura dalam mencari kriteria nasabah macet yang akan memudahkan pemberian kredit untuk calon nasabah repeat order dengan status kelayakannya. Adapun saran dari penelitian ini untuk mengambil keputusan dengan menggunakan metode K-Medoids dengan menggunakan tools orange [5].

Selanjutnya penelitian terkait yang berjudul "Penerapan Data Mining Menggunakan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang" yang membuat kesimpulan bahwa Setelah dilakukan pengelompokkan data mahasiswa dengan menggunakan algoritma *k-means* berdasarkan persebaran kabupaten, asal sekolah, penghasilan ayah dan informasi maka diperoleh suatu gagasan bahwa cluster yang jumlahnya terbanyak adalah strategi promosinya dengan menggunakan brosur dan promosi dari teman atau kerabat dekat.[6]

2.1.1 Landasan teori

2.1.2 Credit Union Sauan Sibarrung

Credit Union Sa'uan Sibarrung adalah sebuah koperasi keuangan mikro yang beroperasi di Indonesia, dengan fokus utama pada pemberdayaan ekonomi masyarakat lokal. Didirikan dengan semangat koperasi, lembaga ini menawarkan berbagai layanan keuangan, termasuk tabungan, pinjaman, dan asuransi, kepada anggotanya. Salah satu ciri khasnya adalah pendekatan berbasis komunitas, di

mana keputusan diambil secara partisipatif oleh para anggota, dengan tujuan untuk meningkatkan kesejahteraan bersama.

Credit Union Sa'uan Sibarrung juga mengutamakan pendidikan keuangan, memberikan pelatihan dan informasi kepada anggotanya untuk meningkatkan pemahaman tentang manajemen keuangan dan perencanaan masa depan. Melalui prinsip-prinsip koperasi dan komitmen terhadap pembangunan berkelanjutan, Credit Union Sauan Sibarrung telah menjadi salah satu kekuatan positif dalam memajukan ekonomi lokal dan meningkatkan inklusi keuangan di wilayahnya.

2.2.2 Data *Mining*

Data mining merupakan suatu proses otomatis terhadap data yang sudah ada. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Data *mining* itu sendiri adalah proses pencarian pola-pola yang tersembunyi (*hidden patern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada didalam database, *werehouse* data, atau media penyimpanan informasi yang lain. Hal penting yang terkait di dalam data mining adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.[7]

Data mining sering juga disebut *Knowledge Discovery in Database* (KDD). Dalam data mining terdapat metode-metode yang dapat digunakan seperti klasifikasi, clustering, regresi, seleksi variabel, dan analisis. Data mining adalah suatu kegiatan analisa data untuk mencari suatu pola tertentu, dengan jumlah data yang besar dan bertujuan utuk menghasilkan informasi yang dapat digunakan dan dikembangkan lebih lanjut. Data mining adalah metode untuk menemukan informasi baru yang berguna dari kumpulan data yang besar dan dapat membantu dalam pengambilan keputusan. Salah satu metode yang bisa digunakan dalam Data Mining adalah metode Clustering yaitu sebuah metode dalam Data Mining yang bersifat tanpa arahan (unsupervised).[8]

2.2.3 *Knowledge Discovery In Database* (KDD)

Knowledge Discovery in Database Process (KDD) adalah salah satu metode yang bisa digunakan dalam melakukan data mining. Fayyed et al. (1996) mendefinisikan KDD sebagai proses dari menggunakan metode data mining untuk mencari informasi-informasi yang berharga, pola yang ada di dalam data, yang melibatkan algoritma untuk mengidentifikasi pola pada data. Dunham (2003) meringkas proses KDD dari berbagai step, yaitu: seleksi data, pra-proses data, transformasi data, data mining, dan yang terakhir interpretasi dan evaluasi. KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2017).[9]

Knowledge Discovery in Database (KDD) adalah cara non-trivial dalam upaya menemukan pola pada sebuah data yang bersifat baru, dapat dimengerti dan

bermanfaat. Pada tahapan KDD terdapat proses data mining, yaitu melakukan analisa dari sejumlah besar kumpulan data observasi.

2.2.4 *Clustering*

Clustering adalah proses pengelompokan benda serupa ke dalam kelompok yang berbeda, atau lebih tepatnya partisi dari sebuah data set kedalam subset, sehingga data dalam setiap subset memiliki arti yang bermanfaat. Algoritma clustering terdiri dari dua bagian yaitu secara hirarkis dan secara partitional. Algoritma hirarkis menemukan cluster secara berurutan dimana klaster ditetapkan sebelumnya, sedangkan algoritma partitional menentukan semua kelompok pada waktu tertentu. Teknik pengelompokan saat ini dapat diklasifikasikan menjadi tiga kategori yaitu partitional, hirarkis dan berbasis lokalitas algoritma. Terdapat satu set objek dan kriteria clustering atau pengelompokan, pengelompokan partitional mememperoleh partisi objek ke dalam cluster sehingga objek dalam klasteralgoritma dari akan lebih mirip dengan benda-benda yang ada di dalam cluster dari pada objek yang terdapat pada klaster yang berbeda.

Diantara beberapa teknik clustering pada data mining yang paling populer adalah K-Means Clustering. Metode *K-Means Clustering* menjadi sangat populer dikarenakan dapat dengat cepat dan efisien mengelompokkan sejumlah besar data dengan pemodelan tanpa supervisi (unsupervised learning). Diantara keunggulan yang dimiliki algoritma *K-Means Clustering* adalah:

a. Relatif mudah diterapkan.

b. Menskalakan ke kumpulan data besar.

c. Menjamin konvergensi.

d. Dapat memperbaharui posisi centroid.

e. Mudah beradaptasi.

f. Menggeneralisasi cluster dengan berbagai bentuk dan ukuran, seperti cluster elips.[10]

2.2.5 *K-Means*

Algoritma *K-means* merupakan salah satu algoritma dengan partitional, karena K-Means didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya. *K-Means clustering* merupakan salah satu metode data clustering non-hirarki yang mengelompokan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil. *K-Means* merupakan salah satu metode pengelompokan data nonhierarki yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok.

Metode *K-Means* hanya akan bekerja pada atribut numerik karena metode ini merupakan algoritma berbasis jarak dari cara kerjanya membagi data menjadi beberapa cluster. Metode K-means merupakan suatu metode yang dapat melakukan pengelompokan data dalam jumlah yang cukup besar dengan perhitungan waktu yang relatif cepat dan efisien. [11]

K-Means adalah salah satu algoritma dari teknik data mining yang mampu melakukan klasterisasi terhadap data heterogen karena pada dasarnya algoritma pengelompokkan hanya mampu mengenali nilai atribut homogen saja.[12]

Langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut:

a. Pilih jumlah *cluster* k.

b. Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberiduberi nilai awal dengan angka-angka random,

c. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana. Untuk menghiutng jarak semua data ke setiap tiitk pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2} \quad \ldots (1)$$

dimana: D (i,j) = Jarak data ke i ke pusat cluster j

Xk i= Data ke i pada atribut data ke k

Xk j= Titik pusat ke j pada atribut ke k

d. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster

tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (*Mean*) bukan satu-satunya ukuran yang bisa dipakai.

e. Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi.[13]

Kelebihan dari metode *K-Means*:

1. Mudah dipahami dan diimplementasikan.
2. Cepat dan efisien untuk data berukuran besar.
3. Cocok untuk data numerik dengan dimensi yang rendah.

Kekurangan dari metode *K-Means*:

1. Membbutuhkan pemilihan nilai k yang tepat.
2. Sensitif terhadap data yang jauh berbeda dari data lainnya.
3. Tidak efektif untuk data dengan bentuk yaang tidak beraturan.

2.2.6 *ELBOW*

Elbow adalah teknik yang digunakan dalam analisis Clustering untuk membantu menentukan jumlah klaster yang optimal dalam suatu dataset. Metode ini disebut "*Elbow*" karena grafik hasil klastering seringkali menyerupai lengkungan siku tangan pada titik di mana penurunan variansi yang dijelaskan oleh klaster tambahan mulai menjadi berkurang secara signifikan.

Proses Metode *Elbow* dimulai dengan mengklasterkan dataset dengan berbagai jumlah klaster yang berbeda. Setiap kali jumlah klaster bertambah, nilai variansi yang dijelaskan oleh klaster meningkat. Setelah itu, nilai variansi yang

dijelaskan oleh setiap klaster diplotkan pada grafik dengan sumbu horizontal yang mewakili jumlah klaster dan sumbu vertikal yang mewakili variansi. Grafik ini akan menunjukkan bagaimana peningkatan jumlah klaster mempengaruhi penurunan variansi. Metode ini dinamakan *Elbow* karena bentuk grafik yang dihasilkan menyerupai siku pada siku tangan manusia[14].

Berdasarkan penelitian yang telah dilakukan sebelumnya dengan judul "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional" Metode Elbow dapat digunakan untuk mencari optimasi dalam penentuan jumlah cluster yang akan diterapkan dalam algoritma clustering, seperti K-Means.

Pada proses perhitungan yang telah dilakukan tentang penerapan clustering dengan menggunakan algoritma K-Means dan metode optimasi cluster yaitu metode Elbow, dapat diambil kesimpulan bahwa penerapan metode Elbow sesuai dengan penelitian ini menghasilkan jumlah cluster terbaik adalah 3, selanjutnya dengan menerapkan metode K-Means dihasilkan data bahwa jumlah siswa sebanyak 66 siswa, dengan proses perhitungan sebanyak 9 iterasi dihasilkan ada 3 cluster yaitu kategori "Siap", "Cukup Siap", dan "Belum Siap", dengan hasil cluster "Siap" sebanyak 7 siswa, cluster "Cukup Siap" sebanyak 30 siswa, dan cluster "Belum Siap" sebanyak 29 siswa.

Hasil tersebut dapat menjadikan pedoman atau pengukuran kekuatan dalam hal jumlah siswa dalam setiap kategorinya untuk pihak sekolah, dengan memberikan tindakan kelas yang sesuai agar jumlah siswa dalam kategori "Siap" menjadi bertambah, dan dalam kategori "Tidak Siap" menjadi berkurang [15].

Penelitian selanjutnya yang berjudul tentang "Metode *Elbow* Dalam Optimasi Jumlah *Cluster* Pada *K-Means Clustering*" menjelaskan bahwa Penentuan titik pusat cluster sangat berpengaruh pada terhadap perhitungan algoritma *k-means*. Dari pengujian menggunakan algoritma *k-means clustering* yang dioptimasikan dengan metode elbow, terbentuk jumlah cluster yang optimal sebanyak 3 cluster. Penentuan k=3 menggunakan metode elbow ini berdasarkan perhitungan selisih dari rata–rata dari tiap cluster yang terbentuk dan nilai Davies Bouldin Index (DBI) yang mendekati 0 menunujukkan cluster yang terbentuk adalah cluster yang optimal. Dalam menentukan jumlah cluster menggunakan metode elbow mendapatkan hasil yang lebih baik tingkat kemiripan setiap anggotanya dibandingkan dengan menentukan jumlah cluster secara acak.

Dari 195 desa dan kelurahan yang berada di Kabupaten Jepara setelah dilakukan clustering, bahwa terdapat 5 desa yang memiliki jumlah penerima Program Keluarga Harapan (PKH) yaitu desa Cepogo, Karanggodang, Lebak Troso dan Tubanan, dimana kelima desa tersebut memiliki jumlah rumah tangga penerima PKH diatas 900 rumah tangga dengan komponen yang paling dominan atau banyak dijumpai adalah dari usia lanjut (lansia) pada penerima PKH di masing-masing desa. Hasil pengelompokkan menggunakan algoritma K-Means Clustering ini menunjukkan bahwa penerima PKH yang mendapatkan prioritas utama berada pada cluster 2. Hasil pengelompokkan ini diharapkan dapat membantu pemerintah dalam menentukan penerima PKH yang akurat dan tepat sasaran sesuai dengan syarat penerima PKH yang telah ditentukan [16].

Beberapa langkah dalam tahap kerja algoritma *Elbow* yaitu:

1. Inisialisasikan pusat *cluster* secara acak sebanyak jumlah *cluster* (k)
2. Setiap data (objek) dialokasikan ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidean Distance* dengan rumus persamaan menurut Prasetyo dalam yaitu:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Keterangan

d : jarak data ke pusat *cluster*

x_i : pusat *cluster* ke-i

y_i : data atribut ke-i

n : banyaknya data

3. Pilih objek pada masing-masing *cluster* secara acak sebagai kandidat medoid baru
4. Hitung jarak setiap objek yang terdapat pada masing-masing *cluster* dengan calon medoid baru
5. Hitung total simpangan (S) dengan menghitung nilai total jarak baru-total jarak lama. Jika didapatkan S < 0, tukarkan objek dengan data *cluster* untuk membuat sekumpulan k objek baru sebagai medoid
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga diperoleh *cluster* serta anggota *cluster* masing-masing.

Kelebihan dalam menggunakan metode Elbow:

1. Penentuan jumlah cluster yang optimal, Salah satu kelebihan utama dari metode Elbow adalah kemampuannya untuk membantu menentukan jumlah klaster yang optimal dalam sebuah dataset. Dengan mengamati grafik "Elbow" yang menampilkan jumlah klaster pada sumbu x dan nilai fungsi objektif (misalnya, nilai inersia atau WCSS - Within-Cluster Sum of Squares) pada sumbu y, kita dapat mengidentifikasi titik di mana penurunan signifikan dalam nilai objektif berkurang secara drastis. Titik ini disebut "siku" (elbow), yang menunjukkan jumlah klaster yang optimal.
2. Simpel dan intuitif, Metode Elbow relatif mudah dipahami dan diimplementasikan. Konsepnya sederhana, yaitu mengamati perubahan nilai objektif saat jumlah klaster meningkat, dan mencari titik di mana penurunan tersebut mulai merata atau menurun secara signifikan.
3. Visualisasi yang jelas, Dengan menggunakan metode Elbow, kita dapat memvisualisasikan hasil analisis klaster secara langsung melalui grafik "Elbow". Ini memberikan pemahaman yang jelas tentang bagaimana nilai objektif berubah seiring dengan peningkatan jumlah klaster, memudahkan interpretasi dan pengambilan keputusan.
4. Pemilihan model yang tepat,Dalam analisis klaster, memilih jumlah klaster yang tepat adalah langkah krusial untuk mendapatkan hasil yang baik. Dengan menggunakan metode Elbow, kita dapat membuat keputusan yang lebih terinformasi tentang jumlah klaster yang optimal, yang pada gilirannya dapat menghasilkan model klaster yang lebih baik.

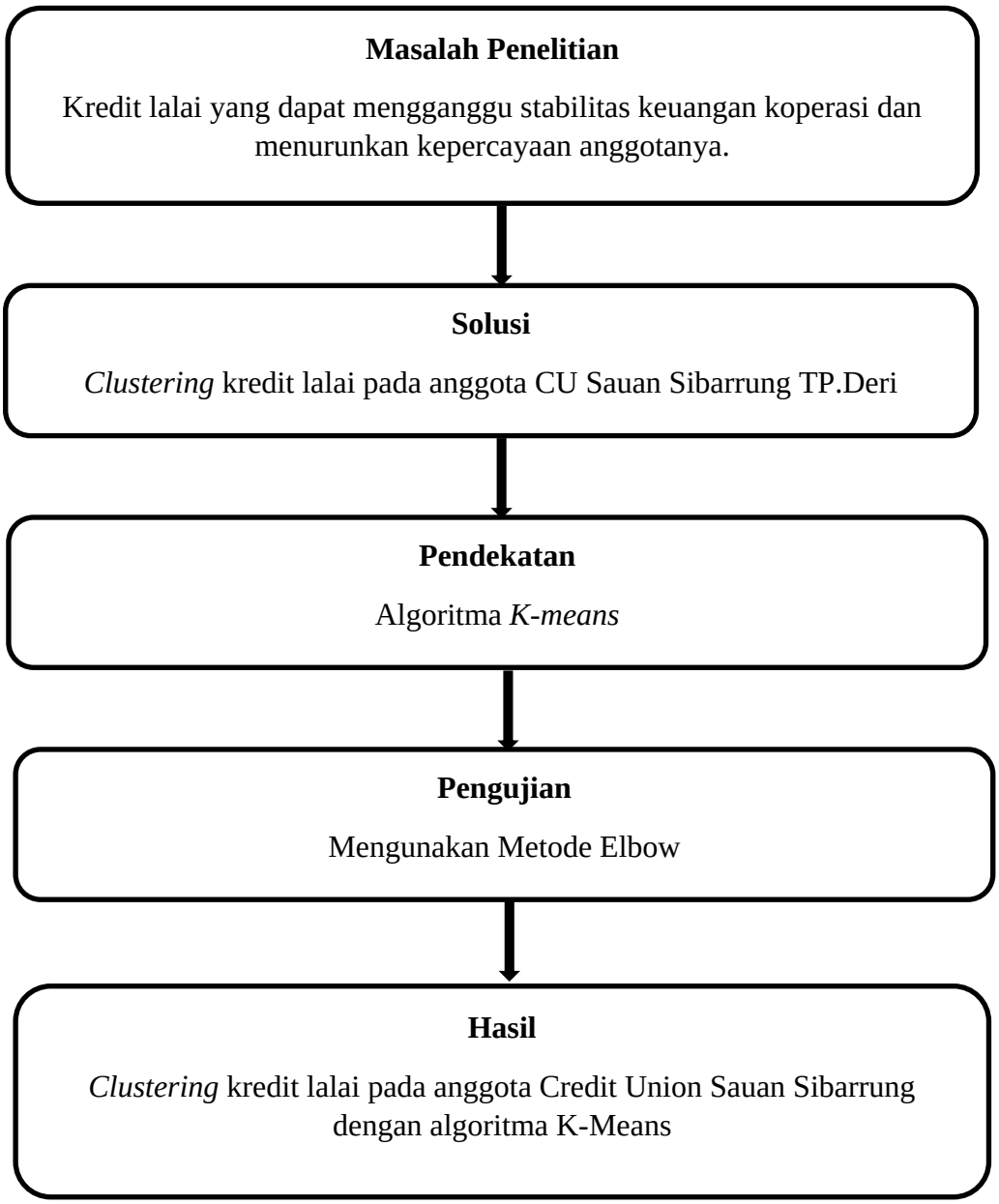
5. Fleksibilitas, Metode Elbow tidak terkait dengan jenis algoritma klaster tertentu. Ini dapat diterapkan pada berbagai metode klastering seperti K-Means, Hierarchical Clustering, dan lainnya. Oleh karena itu, metode ini memiliki fleksibilitas dalam aplikasinya terhadap berbagai jenis dataset dan algoritma klastering.

Kekurang dari metode Elbow:

1. Tidak cocok untuk semua kasus, Ada situasi di mana metode Elbow tidak cocok digunakan, terutama ketika distribusi data tidak homogen atau ketika klaster tidak terpisahkan dengan jelas. Misalnya, dalam kasus data dengan klaster yang berbentuk lingkaran atau tumpang tindih, metode Elbow mungkin tidak dapat mengidentifikasi jumlah klaster yang optimal dengan baik.
2. Hanya Menilai Varians Internal, Metode Elbow hanya mempertimbangkan varians internal dalam data (misalnya, inersia atau WCSS dalam algoritma K-Means), yang mungkin tidak memberikan gambaran yang lengkap tentang kualitas klaster. Varians internal tidak mempertimbangkan apakah klaster tersebut bermakna secara semantik atau apakah klaster tersebut dapat diinterpretasikan dengan baik.
3. Tidak Memberikan Informasi tentang Struktur Klaster, Metode Elbow hanya memberikan informasi tentang jumlah klaster yang optimal, namun tidak memberikan wawasan tentang struktur klaster yang sebenarnya. Misalnya, metode ini tidak memberikan informasi tentang apakah klaster tersebut saling tumpang tindih, kompak, atau terpisah dengan baik.

2.3 Kerangka Pikir

Kerangka pikir penelitian tentang Clustering Nasabah CU dengan metode k-means dapat dilihat pada Gambar 2.1.



Gambar 2.1 Kerangka Pikir