

## BAB II TINJAUAN PUSTAKA

### 2.1. Penelitian Terkait

Penelitian terkait merupakan kumpulan dari temuan penelitian sebelumnya, di mana penelitian sebelumnya digunakan sebagai referensi atau pembandingan dalam penelitian yang sedang dilakukan. Penelitian sebelumnya yang relevan telah menjalankan uji coba Algoritma *K-Medoids* untuk melakukan pemisahan pelanggan. Pusat perhatian penelitian yaitu untuk mengenali jenis dan ciri pelanggan dengan tujuan memberikan saran untuk strategi pemasaran yang lebih efektif dan sesuai. Sumber data yang digunakan berasal dari transaksi penjualan Perum BULOG di kantor wilayah DKI Jakarta & Banten selama tahun 2019, yang mencakup total 9.945 transaksi. Hasil penelitian menunjukkan bahwa jumlah *cluster* optimal adalah 3 (tiga) dengan nilai maksimum *Silhouette Index* sebesar 0,375 dan nilai minimum *Davies Doulbin Index* sebesar 1,030.[5]

Penelitian terkait berikutnya mengenai penerapan Algoritma *K-Medoids* dalam mengklasifikasikan data penyakit pada pasien RSUD Kota Bandung. Penelitian ini bertujuan untuk mendapatkan informasi tentang jenis penyakit yang paling umum dihadapi oleh pasien, dengan menggunakan pendekatan Algoritma *K-Medoids*. Dataset pasien di RSUD Kota Bandung mencakup 8.392 entri. Hasil penelitian menunjukkan bahwa pemilihan 3 *cluster* memberikan hasil optimal, dengan nilai *Silhouette Coefficient* sebesar 0,409373.[6]

Penelitian terkait berikutnya mengkhususkan diri pada analisis Pengelompokan *K-Medoids* dalam konteks penyebaran Covid-19 di Indonesia. Tujuan penelitian ini adalah memahami pola-pola yang muncul dalam menentukan pengelompokan penyebaran Covid-19 di kalangan masyarakat dan mendapatkan set *cluster* yang paling sesuai dengan objek data, sehingga dapat memberikan gambaran yang mendekati realitas dalam pengelompokan data tersebut. Data yang diperoleh 14265 ribu jiwa yang telah terkonfirmasi positif terinfeksi oleh virus Covid-19, sementara sebanyak 2881 jiwa berhasil sembuh dari penyakit ini dan 991 jiwa lainnya telah mengalami kehilangan nyawa. Hasil penelitian dan penerapan menunjukkan bahwa pengelompokan optimal terjadi dengan menggunakan *cluster*.

Dari total 34 data, terdapat 1 data yang tergabung dalam *cluster* pertama, 2 data dalam *cluster* kedua, dan 31 data dalam *cluster* ketiga.[7]

Penelitian lain mengulas pengelompokan data penduduk miskin akibat dampak Covid-19 menggunakan Algoritma *K-Medoids*. Sumber data yang diambil berasal dari Badan Pusat Statistika pada tahun 2020. Pendekatan penelitian ini mengadopsi model CRISP-DM, dengan menggabungkan analisis deskriptif dan pendekatan kuantitatif dalam metodologi penelitiannya. Dengan pendekatan ini, penelitian difokuskan pada hasil analisis data untuk merinci situasi kemiskinan akibat pandemi COVID-19 di Indonesia, dengan tujuan utama mengelompokkan wilayah berdasarkan pola data yang ada. Data hasil pengolahan mengenai tingkat kemiskinan akan disajikan dalam bentuk *cluster*, di mana data tersebut akan dikelompokkan ke dalam dua kategori, yakni tinggi dan rendah. *Cluster* 1 terdiri dari 121 kabupaten/kota dan *Cluster* 2 terdiri dari 427 kabupaten/kota. Evaluasi hasil *clustering* menggunakan *silhouette coefficient* berperan sebagai penilaian kualitas *cluster* yang dihasilkan dari pengujian dataset penduduk miskin tahun 2020. Hasil evaluasi menunjukkan nilai *silhouette coefficient* sebesar 0,4735719. Dengan nilai ini, dapat disimpulkan bahwa *cluster* 2 dianggap sebagai *cluster* optimal karena mendekati nilai 1.[8]

Penelitian berbeda dengan penelitian sebelumnya, perbedaannya yaitu pada jenis kasus. Hal yang membedakan lainnya yaitu kualitas *cluster* yang dihasilkan pada penelitian terdahulu lebih rendah. Hal yang membedakan lainnya yaitu pada jumlah *cluster*, dimana penelitian ini terdapat 2 *cluster* sedangkan penelitian sebelumnya kebanyakan menggunakan 3 *cluster*. Data diambil langsung dari Balai Desa Kalierang melalui wawancara dengan masyarakat Kalierang.

## **2.2.Landasan Teori**

### **2.2.1. K-Medoids**

*K-Medoids* merupakan algoritma yang digunakan untuk mengidentifikasi *medoids* dalam sebuah kelompok (*cluster*), yang merupakan titik pusat yang merepresentasikan kelompok tersebut. Perbedaan utama antara *K-Medoids* dan *K-Means* terletak pada pendekatan pemilihan representatif kelompok. Pada algoritma *K-Medoids*, kita mencari objek data  $k$  yang menjadi *medoid* sehingga jumlah ketidaksamaan antar objek data dalam kelompok

diminimalkan. Dari perspektif lain, K-Means memanfaatkan total jarak Euclidean untuk menetapkan pusat objek dalam kelompok.

Metode *K-Medoids* dianggap lebih unggul karena pendekatannya yang lebih robust dalam menentukan objek yang mewakili kelompok. Dengan mencari *medoids*, algoritma ini lebih tahan terhadap nilai *outlier* atau anomali yang dapat mempengaruhi hasil *clustering*. Di lain pihak, *K-Means* dapat dipengaruhi oleh data yang berada di luar kelompok utama karena mengandalkan total jarak *euclidean*.

Dengan fokus pada kualitas representasi kelompok, *K-Medoids* membantu meningkatkan akurasi dan keandalan hasil *clustering*, sehingga banyak peneliti dan praktisi lebih memilih algoritma ini untuk analisis data yang lebih efektif dan informatif. Langkah-langkah dalam algoritma ini melibatkan beberapa tahapan, di antaranya:

1. Memulai dengan menginisialisasi pusat *cluster* sebanyak  $k$  (jumlah *cluster*).
2. Mengalokasikan setiap data (objek) ke *cluster* terdekat menggunakan rumus jarak *Euclidean Distance* yang dihitung dengan persamaan:  $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$  (2).
3. Secara acak, memilih objek pada setiap *cluster* sebagai kandidat *medoids* baru.
4. Menghitung jarak setiap objek dalam setiap *cluster* dengan kandidat *medoids* baru.
5. Menghitung total simpangan (S) dengan mengurangkan nilai total *distance* baru dengan total *distance* lama. Jika  $S < 0$ , dilakukan pertukaran objek dengan data *cluster* untuk membentuk kumpulan  $k$  objek baru sebagai *medoid*.
6. Mengulangi langkah 3 sampai 5 hingga tidak ada perubahan *medoids*, sehingga diperoleh *cluster* beserta anggota masing-masing *cluster*. [7]

### 2.2.2. Clustering(pengelompokan)

Proses *clustering* merupakan tahap yang sangat signifikan dalam proses *data mining*, dimana tujuannya adalah untuk mengelompokkan data atau objek-objek yang memiliki karakteristik yang serupa. Dalam konteks *clustering*, tidak ada variabel target yang telah ditentukan sebelumnya, berbeda

dengan pendekatan klasifikasi. Berbagai algoritma *clustering*, seperti *K-Means*, *Improved K-Means*, *Fuzzy C-Means*, *DBSCAN*, *K-Medoids (PAM)*, *CLARANS*, dan *Fuzzy Subtractive*, telah digunakan oleh para peneliti. Meskipun setiap algoritma memiliki kelebihan dan kelemahan unik, prinsip dasarnya tetap sama, yaitu mengelompokkan data berdasarkan kesamaan karakteristik dan mengukur jarak kemiripan antar data di setiap kelompok. *Clustering* sering menjadi langkah awal yang penting dalam proses analisis data, karena dapat membantu mengidentifikasi pola-pola tersembunyi atau kelompok-kelompok yang serupa dalam data. Hal ini memungkinkan para peneliti atau analis data untuk memahami lebih dalam tentang data yang mereka miliki dan membuat keputusan yang lebih informasional.

Pemilihan algoritma *clustering* yang tepat harus didasarkan pada karakteristik data yang akan diolah dan tujuan analisis yang ingin dicapai. Misalnya, *K-Means* cocok untuk data yang memiliki bentuk kelompok yang jelas dan *non-hierarchical*, sedangkan *DBSCAN* lebih efektif dalam mengidentifikasi kelompok yang berbentuk tidak teratur dan bisa menangani *noise* dalam data. Meskipun ada banyak algoritma *clustering* yang telah digunakan dan dikembangkan, tetapi semuanya berusaha untuk mencapai tujuan yang sama, yaitu mengelompokkan data berdasarkan kesamaan karakteristik, sehingga memudahkan pengguna untuk mendapatkan wawasan baru dan informasi berharga dari data yang ada.[7]

### **2.2.3. Silhouette Coefficient**

Salah satu metode analisis untuk mengukur validitas dalam proses *clustering* adalah menggunakan *Silhouette Coefficient Index*. Rentang nilai *Silhouette Coefficient* berkisar antara -1 hingga 1. Jika nilai  $si = 1$ , ini menunjukkan bahwa objek  $i$  telah ditempatkan dengan baik dalam *cluster* yang sesuai. Sebaliknya, jika nilai  $si = 0$ , maka posisi objek  $i$  terletak di antara dua *cluster*, sehingga penempatannya tidak dapat dipastikan, apakah harus dalam *cluster* A atau B. Sementara jika nilai  $si = -1$ , hal ini mengindikasikan adanya tumpang tindih dalam struktur *cluster* yang dihasilkan, sehingga objek  $i$  sebaiknya ditempatkan dalam *cluster* yang berbeda.

Melalui perhitungan *Silhouette Coefficient*, kita dapat memperoleh informasi tentang seberapa baik setiap objek ditempatkan dalam *cluster* dan seberapa jelas perbedaan antara *cluster*. Metode ini membantu dalam mengevaluasi kualitas dan akurasi hasil dari proses *clustering*, serta membantu dalam memahami apakah objek-objek sudah dikelompokkan dengan tepat atau apakah terdapat kebingungan dalam penempatannya.

Dengan menggunakan *Silhouette Coefficient*, kita dapat mendapatkan wawasan yang lebih mendalam tentang efektivitas dari metode *clustering* yang digunakan, dan hal ini membantu dalam pemilihan jumlah *cluster* yang optimal serta pemahaman tentang kesesuaian model *clustering* dengan data yang ada.[9]

#### **2.2.4. Data Mining**

*Data mining* merupakan suatu pendekatan yang digunakan untuk menggali informasi dari basis data yang memiliki ukuran besar dan kompleks, dengan tujuan untuk mengidentifikasi pola atau wawasan yang bernilai. Proses *data mining* melibatkan penerapan berbagai teknik dan algoritma untuk menemukan pola tersembunyi dalam data, sehingga dapat memberikan dasar untuk pengambilan keputusan yang lebih cerdas dan memperoleh informasi yang lebih signifikan. Dalam *data mining*, terdapat empat kelompok utama dalam pengelompokan kasus, yaitu:

1. Model prediksi (*prediction modelling*): Pusat perhatian ditujukan pada pengembangan model yang mampu meramalkan nilai atau peristiwa yang akan terjadi di masa yang akan datang berdasarkan data historis yang telah ada. Model ini dapat membantu dalam meramalkan tren, melakukan prediksi bisnis, dan membuat keputusan berdasarkan analisis data yang mendalam.
2. Analisis *cluster* (*cluster analysis*): Mengelompokkan objek-objek data menjadi beberapa kelompok atau *cluster* berdasarkan kemiripan atau kesamaan karakteristik tertentu. Hal ini membantu dalam mengidentifikasi kategori atau kelompok yang serupa, sehingga memudahkan dalam pemahaman dan pengelompokan data.

3. Analisis asosiasi (*association analysis*): Bertujuan untuk menemukan hubungan atau keterkaitan antara berbagai item atau atribut dalam data. Dengan analisis asosiasi, kita dapat mengidentifikasi pola pembelian bersama atau korelasi di antara item-item tertentu.
4. Deteksi anomali (*anomaly detection*): Fokus pada identifikasi objek atau data yang dianggap tidak biasa atau berbeda dari mayoritas data. Hal ini membantu dalam menemukan potensi kejadian yang tidak wajar atau masalah dalam data yang perlu ditinjau lebih lanjut.

Menerapkan *data mining* pada berbagai bidang memberikan manfaat dalam mengungkap informasi berharga dari data yang besar dan kompleks. Dengan memahami berbagai kelompok analisis data mining ini, kita dapat mengoptimalkan analisis data dan pengambilan keputusan dalam berbagai konteks dan industri.[1]

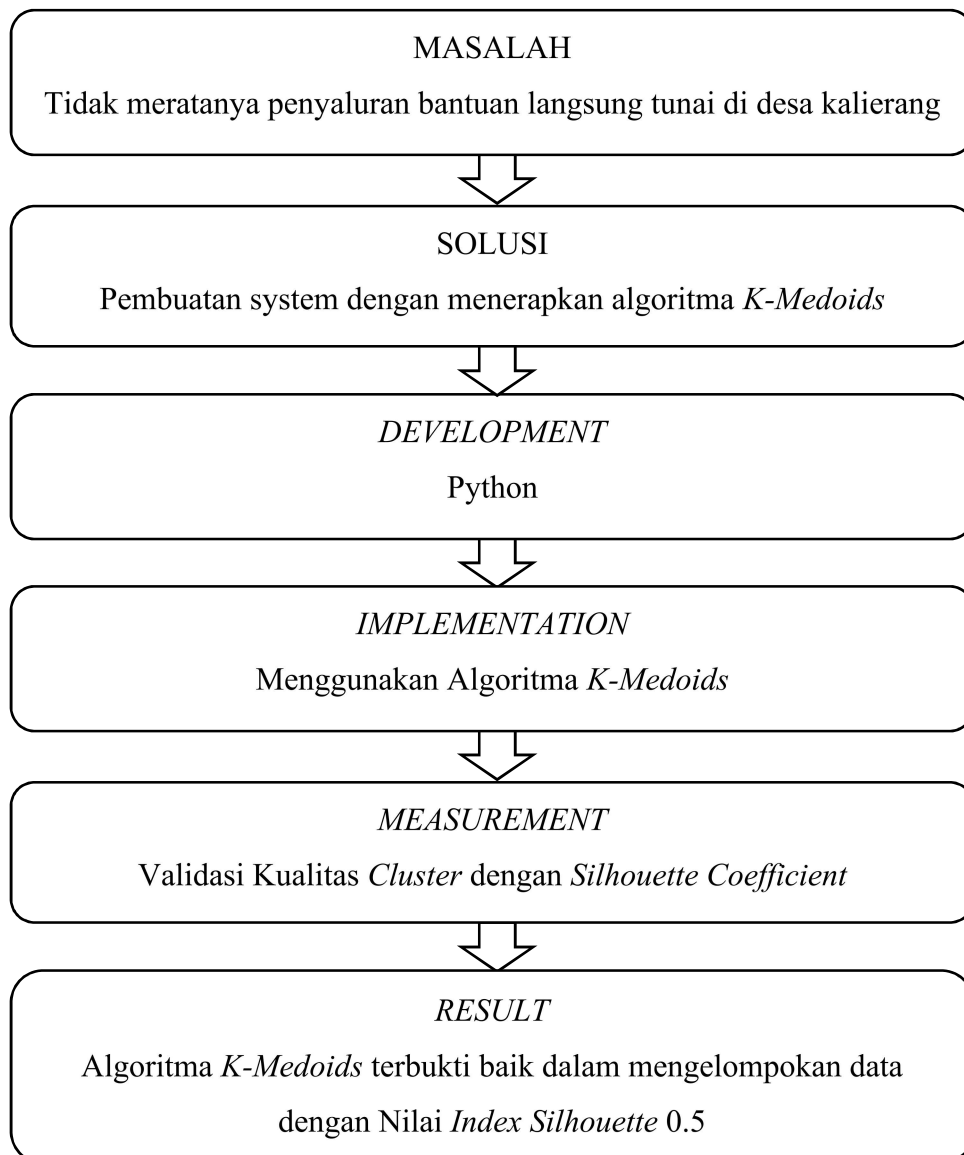
*Data mining* memiliki peran krusial dalam mengenali pola atau informasi berharga dari data yang telah dipilih. Proses ini melibatkan penerapan beragam teknik, metode, dan algoritma. Pemilihan metode atau algoritma yang cocok dengan tujuan analisis dan keseluruhan proses *Knowledge Discovery in Databases* (KDD) menjadi hal yang sangat penting. Dalam praktik *data mining*, ada banyak pilihan teknik dan algoritma yang dapat digunakan untuk menganalisis data. Pemilihan metode yang tepat harus didasarkan pada kebutuhan spesifik dari tugas analisis dan tujuan dari proses KDD secara keseluruhan. Selain itu, karakteristik data yang akan diolah juga harus diperhatikan dalam memilih teknik yang paling sesuai. Memahami variasi teknik dan algoritma dalam *data mining* membantu dalam mencapai hasil analisis yang akurat dan relevan. Dengan pendekatan yang tepat, *data mining* dapat menjadi alat yang kuat untuk mengungkapkan wawasan berharga dan pola-pola tersembunyi dari data yang ada. Oleh karena itu, memilih metode atau algoritma yang sesuai dengan konteks dan tujuan analisis menjadi kunci keberhasilan dalam proses *data mining*.[2]

### 2.2.5. Python

Guido Van Rossum mengembangkan bahasa pemrograman Python pada tahun 1989 di CWI Amsterdam. Python masuk dalam *kategori interpreted language* dan memiliki dasar yang berasal dari bahasa pemrograman C++. Keunggulan Python sebagai bahasa pemrograman yang bersifat *open source* membuatnya dapat didistribusikan secara gratis, menjadikannya pilihan yang populer untuk berbagai kebutuhan pengembangan perangkat lunak di berbagai sistem informasi. Python dapat digunakan untuk mengembangkan aplikasi hampir di semua platform, termasuk untuk aplikasi *desktop*, *website*, dan perangkat *mobile*. Keunggulan Python semakin diperkuat dengan kemajuan teknologi kecerdasan buatan (*Artificial Intelligence*), karena bahasa ini memiliki dukungan modul-modul yang banyak digunakan dalam penelitian, seperti *machine learning* dan *deep learning*. [10] Bahasa pemrograman Python menonjol dengan berbagai fasilitasnya, termasuk kelimpahan *library* dan *modul* yang sangat umum digunakan dalam pengembangan perangkat lunak. Kemudahan dalam tata bahasa membuatnya mudah dipelajari, sementara dukungan terhadap paradigma pemrograman berorientasi objek (OOP) meningkatkan fleksibilitas. Python juga dikenal dengan aturan *layout* kode sumbernya yang mempermudah dalam proses pengecekan dan *debugging*. Banyaknya keunggulan yang dimiliki, Python menjadi salah satu bahasa pemrograman yang paling populer dan banyak digunakan oleh para pengembang perangkat lunak untuk berbagai keperluan, termasuk dalam bidang ilmu pengetahuan seperti kecerdasan buatan (AI) dan *machine learning*.

### 2.2.6. Kerangka Pemikiran

Dari latar belakang yang telah dijelaskan sebelumnya, penulis merasa terdapat permasalahan yang mendorong untuk membangun sebuah model komputasi dengan menggunakan metode *K-Medoids* untuk pengelompokan data penerima bantuan langsung tunai. Penelitian ini mengadopsi kerangka pemikiran yang terlihat pada gambar 2.1 sebagai landasan dalam menjalankan penelitian tersebut.



**Gambar 2. 1** Kerangka Berpikir

Gambar 2.1 merupakan kerangka penelitian yang dilakukan dalam melakukan pengelompokan data penerima BLT dana Desa Kalierang. Penjelasan dari tiap-tiap tahap pada kerangka penelitian akan dijelaskan lebih lanjut yaitu sebagai berikut:

#### 1. MASALAH

Penelitian ini bertujuan untuk mengatasi beberapa permasalahan yang muncul dalam penyaluran bantuan langsung tunai di desa Kalierang. Tidak meratanya pembagian dana BLT menyebabkan ketidakpuasan di kalangan warga, karena masih banyak rumah tangga miskin yang belum terdaftar sebagai penerima. Selain itu, terdapat kesalahan dalam penetapan sasaran penerima BLT, di mana beberapa



orang yang seharusnya tidak memenuhi syarat juga menerima bantuan tersebut. Contohnya, beberapa anggota POLRI, ASN, dan perangkat desa yang sebenarnya mampu menerima BLT, tetapi masih mendapatkannya. Selain itu, ada juga kasus di mana warga yang memiliki kendaraan mewah atau perhiasan emas juga mendapatkan BLT, meskipun seharusnya mereka tidak memenuhi kriteria penerima. Masalah ini diperparah oleh data ganda, di mana sekitar 15% warga terdata ganda dan menerima bantuan ganda. Dengan melihat kondisi tersebut, model komputasi menggunakan metode *K-Medoids* diusulkan untuk memperbaiki pengelompokan data penerima BLT dan mengatasi permasalahan tersebut.

## 2. SOLUSI

Untuk mengatasi masalah yang telah diuraikan sebelumnya, solusinya adalah dengan mengimplementasikan sebuah sistem yang menggunakan metode *K-Medoids* untuk melakukan pengelompokan data penerima bantuan langsung tunai. Pemanfaatan teknologi informasi, terutama dengan memanfaatkan algoritma *K-Medoids* yang telah banyak digunakan dalam penelitian sebelumnya, diharapkan dapat memberikan dampak besar dalam memperbaiki aktivitas manusia terkait penyaluran bantuan. Dengan menerapkan metode *K-Medoids*, sistem akan melakukan pengelompokan data penerima bantuan berdasarkan kesamaan atau keterkaitan tertentu. Hal ini diharapkan dapat meningkatkan efisiensi dan keakuratan dalam penentuan penerima bantuan langsung tunai. Penggunaan algoritma *K-Medoids* dalam klustering telah terbukti efektif dan cocok untuk proses pengelompokan data seperti ini. Dengan adanya sistem yang menggunakan metode *K-Medoids*, diharapkan penyaluran dana BLT akan menjadi lebih merata dan tepat sasaran. Dampak negatif seperti penerima ganda dan penerima yang tidak memenuhi syarat dapat diminimalisir dengan adanya pengelompokan yang lebih akurat. Dengan begitu, keadilan dan kepuasan warga terkait hasil dari penyaluran bantuan dapat ditingkatkan.

### 3. *DEVELOPMENT*

Pengembangan sistem dalam penelitian ini menggunakan Bahasa pemrograman python.

### 4. *IMPLEMENTATION*

Implementasi sistem pengelompokan data penerima blt dengan menerapkan metode *K-Medoids* dimulai dari pengumpulan data, normalisasi data dan proses klusterisasi data, dimana akan dijabarkan lebih rinci di bab 3

### 5. *MEASUREMENT*

Dalam implementasi metode *K-Medoids*, ada suatu ukuran yang akan diambil sebagai acuan untuk mengukur kualitasnya, yaitu nilai *Silhouette Coefficient*.

### 6. *RESULT*

Pencapaian dari penelitian ini adalah bahwa metode *K-Medoids* dapat digunakan untuk mengelompokkan data penerima BLT.