

BAB II KERANGKA TEORITIS

A. Landasan Teori

1. Algoritma K-Means

Untuk melakukan pemetaan daerah, terdapat beberapa metode klusterisasi yang dapat digunakan. (Arhami & Nasir, 2020, p. 148) menjelaskan bahwa algoritma K-Means adalah salah satu algoritma pengklusteran yang mengelompokkan data dengan melihat pada keserupaan karakteristik ataupun ciri-ciri tiap objek data sehingga menghasilkan kelompok yang disebut kluster; algoritma K-Means mudah untuk diimplementasikan karena kompleksitas waktu dan ruang yang relatif kecil sehingga memberikan efisiensi dalam komputasinya hingga dapat menghasilkan dengan cukup baik.

Menurut (Arhami & Nasir, 2020, p. 148), langkah-langkah dasar untuk algoritma K-Means adalah;

- tentukan jumlah k *cluster* sesuai dengan kebutuhan;
- membagi anggota *cluster awal* secara acak;
- menentukan titik pusat awal atau biasa disebut centroid dengan menggunakan rumus berikut;

$$M_k = \left(\frac{1}{n_k}\right) \sum_{i=1}^{n_k} X_{ik} \quad (2.1)$$

- menghitung *square error* pada setiap kluster C_k yang merupakan jumlah kuadrat dari jarak Euclidean antara tiap sampel dalam C_k dan titik pusat (*centroid*); error ini juga dikenal dengan nama *within cluster variation* (WCV), yaitu;

$$e_k^2 = \sum_{i=1}^{n_k} (X_{ik} - M_k)^2 \quad (2.2)$$

- kemudian menghitung jumlah dari keseluruhan error pada k-kluster dengan rumus;

$$E_k^2 = \sum_{k=1}^k e_k^2 \quad (2.3)$$

- Kelompokkan kembali semua sampel berdasarkan jarak minimumnya dari setiap pusat, berikan distribusi sampel baru sesuai dengan clusternya; Untuk mendapatkan distribusi sampel baru, dapat dilakukan dengan menghitung jarak antara setiap titik pusat dan seluruh sampel; perhitungan jarak dapat menggunakan beberapa metode, contohnya dengan *Euclidean Distance*;

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2.4)$$

- (g) Catat hasil untuk anggota klaster baru sesuai dengan hasil yang diperoleh pada langkah 5; Setelah mengulangi langkah 3 beberapa kali, dapat terlihat bahwa nilai square error berkurang secara signifikan nanti.

Menurut (Arhami & Nasir, 2020), perhitungan algoritma K-Means dapat dijabarkan dengan contoh data sebagai berikut;

| x_k | p | q |
|-------|----------|----------|
| x_1 | 0 | 2 |
| x_2 | 0 | 0 |
| x_3 | 1.5 | 0 |
| x_4 | 5 | 0 |
| x_5 | 5 | 2 |

- (a) misalnya untuk kasus ini, nilai K yang diambil adalah $K = 2$, maka *cluster* yang diambil secara acak adalah sebagai berikut;

$$C_1 = \{x_1, x_2, x_4\}$$

$$C_2 = \{x_3, x_5\}$$

- (b) maka dapat ditentukan titik tengah dari kedua *cluster* tersebut, yaitu;

$$M_1 = \left\{ \left(\frac{0 + 0 + 5}{3} \right), \left(\frac{2 + 0 + 0}{3} \right) \right\} = \{1.66, 0.66\}$$

$$M_2 = \left\{ \left(\frac{1.5 + 5}{2} \right), \left(\frac{0 + 2}{2} \right) \right\} = \{3.25, 1.00\}$$

- (c) *square error* diperoleh adalah;

$$e_1^2 = [(0 - 1.66)^2 + (2 - 0.66)^2 + [(0 - 1.66)^2 + (0 - 0.66)^2] + [(5 - 1.66)^2 + (0 - 0.66)^2]] = 19.36$$

$$e_2^2 = [(1.5 - 3.25)^2 + (0 - 1)^2 + [(5 - 3.25)^2 + (2 - 1)^2]] = 8.12$$

Sehingga total *square error*-nya adalah;

$$E^2 = e_1^2 + e_2^2 = 19.36 + 8.12 = 27.48$$

- (d) berikutnya adalah dilakukan redistribusi sampel ke dalam kelas yang sesuai dengan hasil perhitungan jarak terpendek yang dimiliki oleh tiap sampel dengan jarak minimum dari titik pusat *cluster* yang ada akan menjadi *cluster* baru bagi sampelnya;

$$d(M_1, x_1) = \sqrt{(1.66 - 0)^2 + (0.66 - 2)^2} = 2.14$$

$$d(M_2, x_1) = \sqrt{(3.25 - 0)^2 + (1 - 2)^2} = 3.40$$

$$d(M_1, x_2) = \sqrt{(1.66 - 0)^2 + (0.66 - 0)^2} = 1.79$$

$$d(M_2, x_2) = \sqrt{(3.25 - 0)^2 + (1 - 0)^2} = 3.40$$

$$d(M_1, x_3) = \sqrt{(1.66 - 1.5)^2 + (0.66 - 0)^2} = 0.83$$

$$d(M_2, x_3) = \sqrt{(3.25 - 1.5)^2 + (1 - 0)^2} = 2.01$$

$$d(M_1, x_4) = \sqrt{(1.66 - 5)^2 + (0.66 - 0)^2} = 3.41$$

$$d(M_2, x_4) = \sqrt{(3.25 - 5)^2 + (1 - 0)^2} = 2.01$$

$$d(M_1, x_5) = \sqrt{(1.66 - 5)^2 + (0.66 - 2)^2} = 3.60$$

$$d(M_2, x_5) = \sqrt{(3.25 - 5)^2 + (1 - 2)^2} = 2.01$$

untuk memudahkan penentuan klasternya maka hasil perhitungan diatas dapat dibuat dalam bentuk tabel seperti berikut ini dan jarak minimum dari titik tengah *cluster* ke sampel menjadi kelas sampel;

| x_k | $d(M_1, x_{k2})$ | $d(M_2, x_{k2})$ | Klaster |
|-------|------------------|------------------|---------|
| x_1 | 2.14 | 3.40 | C_1 |
| x_2 | 1.79 | 3.40 | C_1 |
| x_3 | 0.83 | 2.01 | C_1 |
| x_4 | 3.41 | 2.01 | C_2 |
| x_5 | 3.60 | 2.01 | C_2 |

sehingga *cluster* baru dapat dituliskan dengan $C_1 = \{x_1, x_2, x_3\}, C_2 = \{x_4, x_5\}$; kedua *cluster* tersebut dihitung kembali titik tengah dan total *square error*-nya untuk memastikan apakah perlu dilanjutkan ke iterasi berikutnya atau tidak; titik tengah dari kedua *cluster* tersebut adalah;

$$M_1 = \left\{ \left(\frac{0 + 0 + 1.5}{3} \right), \left(\frac{2 + 0 + 0}{3} \right) \right\} = \{0.5, 0.66\}$$

$$M_2 = \left\{ \left(\frac{5 + 5}{2} \right), \left(\frac{0 + 2}{2} \right) \right\} = \{5, 1.00\}$$

dan *square error*-nya adalah: $e_1^2 = 4.17$ dan $e_2^2 = 2$ sehingga totalnya diperoleh;

$$E^2 = e_1^2 + e_2^2 = 4.17 + 2 = 6.17$$

jika dilihat total *square error*-nya yang turun begitu signifikan dari 27.48 ke 6.17, maka untuk kasus sederhana ini iterasi pertama sudah cukup untuk mengambil kesimpulan bahwa *cluster* yang dihasilkan adalah $C_1 = \{x_1, x_2, x_3\}, C_2 = \{x_4, x_5\}$.

2. *Data mining dan Clustering*

Untuk menemukan suatu informasi pada data, maka dibutuhkan suatu metode. Menurut (Arhami & Nasir, 2020, p. 2), Data mining adalah teknik untuk mengekstraksi informasi dan pola yang berguna dari sejumlah besar data; data mining meliputi pengumpulan data, ekstraksi data, analisis data, statistik data menggunakan analisis matematis untuk mendapatkan atau mendeteksi pola atau tren dari data; pola ini sulit ditemukan oleh eksplorasi data secara biasa/tradisional karena terlalu rumitnya hubungan antar data atau juga dapat disebabkan oleh data yang begitu besar; *data mining* memiliki berbagai metode sehingga secara sederhana *data mining* dapat digambarkan sebagai suatu pola atau model atau kaidah atau pengetahuan yang dihasilkan dari *data mining*. Mengacu pada penjelasan tersebut, dengan *data mining* akan didapatkan informasi dalam data yang begitu besar melalui pengolahan yang tepat dan efektif.

(Arhami & Nasir, 2020, p. 4) menyatakan bahwa *data mining* biasanya dilakukan melalui tiga langkah, yaitu (1) eksplorasi, (2) pemodelan atau identifikasi pola, (3) penerapan; pada langkah pertama dilakukan persiapan data dalam jumlah besar, kemudian data dibersihkan atau direduksi sesuai kebutuhan, lalu menghilangkan data duplikat sehingga data yang tersisa adalah data yang siap untuk digunakan; kemudian, pada langkah kedua, dilakukan dengan membangun dan mengevaluasi model statistik sehingga dapat membuat prediksi terbaik dan paling optimal; pada langkah terakhir, model yang digunakan diuji pada data pelatihan dan uji untuk menghasilkan prediksi atau perkiraan hasil yang diharapkan.

Untuk melakukan pengelompokan, dibutuhkan suatu teknik yang tepat sesuai dengan hasil yang diharapkan. Menurut (Jollyta, Ramdhan, & Zarlis, 2020, p. 53), Pengklasteran adalah model di mana populasi heterogen dibagi menjadi serangkaian klaster homogen; pengklasteran dilakukan dengan mengumpulkan data clustered dari kumpulan data yang lebih besar; pengklasteran menggabungkan kelompok minoritas yang tersebar menjadi kelompok yang lebih besar dengan entitas yang serupa; pengklasteran dapat digunakan untuk secara otomatis menemukan klaster dalam kumpulan data yang berdekatan yang memiliki arti khusus di semua variabel. terdapat 4 pendekatan untuk permasalahan *clustering*, yaitu, *partition clustering*, *grid – base clustering*, *hierarchical clustering*, dan *density- based clustering*. Pada penelitian ini, model *clustering* digunakan untuk mengelompokan data daerah rawan tindak kekerasan terhadap perempuan dan anak, dengan pendekatan *partition clustering*.

3. Model Data Flat-File

Komponen terpenting dalam sistem informasi adalah data. Menurut (Bernard, 2012, p. 130) , data adalah fakta kasar tentang orang, tempat, peristiwa, dan hal-hal yang dapat disusun menjadi kumpulan fakta dan angka yang dapat diolah menjadi informasi yang berguna. Oleh sebab itu, dalam penelitian ini data yang digunakan merupakan data fakta dari kejadian yang sebenarnya terjadi.

(Simarmata, Haviluddin, Ahmar, & Hidayat, 2018, p. 125) menyatakan bahwa model data flat file sebagai satu file dengan format baris dan kolom yang sederhana sehingga semua data disimpan dalam satu file; Informasi pada suatu flat-file disimpan sebagai fields, dengan fields-nya yang dipisahkan menjadi beberapa karakter (delimiter). Dari penjelasan tersebut maka data per-individual tidak terkait dengan file lainnya.

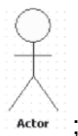
4. Unified Modeling Language (UML)

Dalam perancangan sistem, dibutuhkan suatu pemodelan yang dapat menggambarkan sistem. (Henderi, *Object Oriented Modelling With Unified Modeling Language (UML)*, 2007, p. 4) menyatakan bahwa *Unified Modeling Language (UML)* adalah bahasa pemodelan yang telah menjadi standar industri perangkat lunak untuk memvisualisasikan, merancang, dan mendokumentasikan sistem perangkat lunak; sehingga UML biasa digunakan untuk menggambarkan batasan sistem, proses bisnis, struktur statik suatu sistem, sifat suatu sistem, arsitektur implementasi dan memperluas *fungsi*lity.

Menurut (Kusnadi, Supiandi, Syabaniah, & Oktapiani, 2020), UML memiliki berbagai diagram yang dapat mewakili berbagai sudut pandang dari sistem/ perangkat lunak yang akan dibangun, yaitu;

(a) *usecase* diagram;



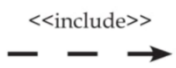

menyatakan bahwa *usecase* diagram mendeskripsikan siapa yang akan menggunakan sistem dan cara pengguna berinteraksi dengan sistem sedangkan *usecase* naratif menggambarkan langkah-langkah tiap interaksi secara tekstual; berikut adalah simbol pada *usecase*;



actor menggambarkan seseorang yang berinteraksi dengan sistem;



usecase menggambarkan apa yang dikerjakan sistem, biasanya menggunakan kata kerja;


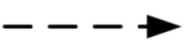

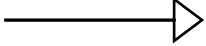
| | | |
|---|---|--|
|  | ; | <i>association</i> digunakan untuk menghubungkan antara aktor dengan <i>usecase</i> ; |
|  | ; | <i>association</i> digunakan untuk menghubungkan aktor yang berinteraksi secara pasif dengan sistem; |
|  | ; | <i>include</i> digunakan untuk menghubungkan <i>usecase</i> dengan <i>usecase</i> yang prosesnya diharuskan atau menjadi syarat yang harus dipenuhi; |
|  | ; | <i>extend</i> digunakan sebagai perluasan dari <i>usecase</i> lain jika kondisi terpenuhi; |

(b) *class diagram*;

class diagram menggambarkan struktur objek sistem dengan menunjukkan kelas objek yang membentuk sistem dan hubungan antara kelas objek tersebut. Di bawah ini adalah contoh *class diagram*;

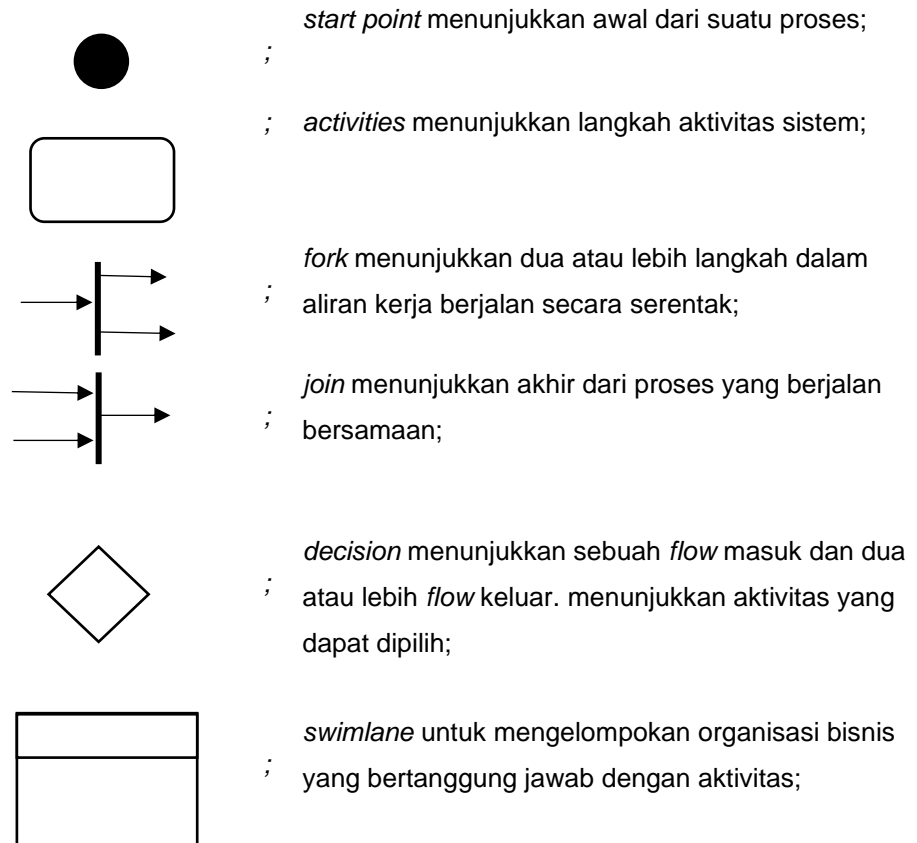
| | |
|---|-------------------|
| <i>Subscription</i> | <i>Class name</i> |
| <i>Series: String</i> | <i>Attributes</i> |
| <i>PriceCategory: Ctegrory</i> | |
| <i>Number: Integer</i> | |
| <i>Cost(): Money</i> | <i>Operation</i> |
| <i>Reserve (Series: String Leve;: SeatLevelCancel()</i> | |

berikut adalah notasi *class diagram*;

| | | |
|---|---|---|
|  | ; | <i>association</i> merupakan deskripsi koneksi diantara <i>instance</i> kelas; |
|  | ; | <i>dependency</i> untuk menghubungkan antara dua elemen model; |
|  | ; | <i>realization</i> untuk menghubungkan antara spesifikasi dan implementasinya; |
|  | ; | <i>generalization</i> untuk menghubungkan antara deskripsi yang lebih spesifik dan lebih umum, digunakan untuk pewarisan deklarasi tipe polimorfik; |
| «kind» | ; | <i>usage</i> digunakan jika ada pada situasi dimana satu elemen memerlukan elemen lain agar berfungsi dengan benar; |

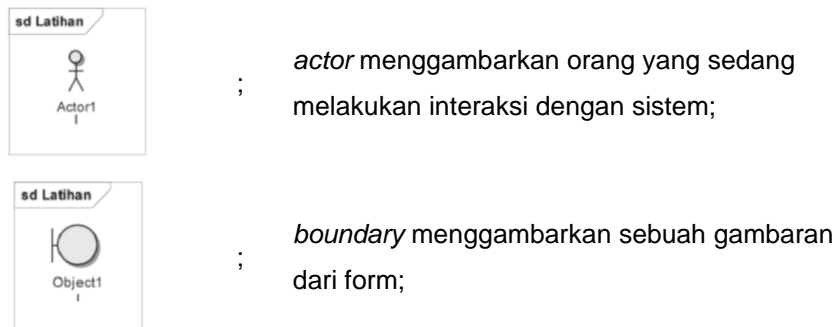
(c) *activity* diagram;

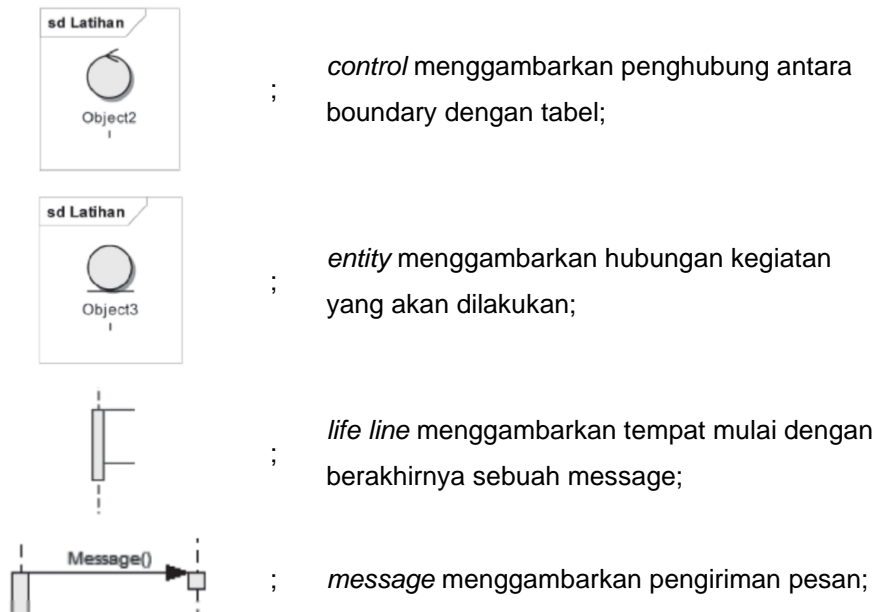
activity diagram digunakan untuk menggambarkan arus aktivitas, memodelkan tindakan yang dilakukan saat operasi dijalankan, dan memodelkan hasil dari tindakan tersebut. Simbol *activity* diagram adalah:



(d) *sequence* diagram;

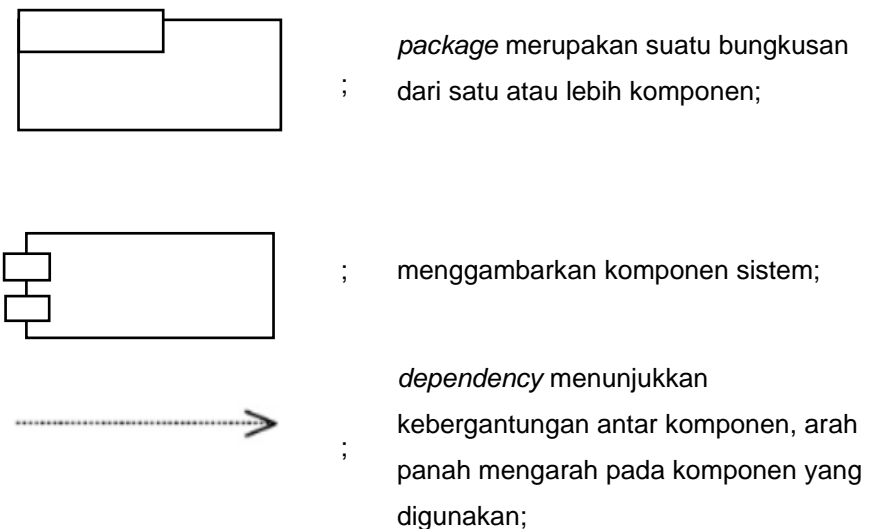
sequence diagram menggambarkan bagaimana objek berinteraksi antara satu sama lain melalui pesan sekuensi dengan mengilustrasikan pesan terkirim dan diterima antara objek dalam sekuensi dan *timing*-nya; berikut adalah simbol pada *sequence* diagram;





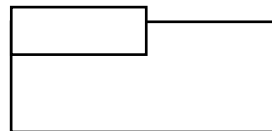
(e) *component* diagram;

component diagram digunakan untuk menunjukkan bagaimana kode pemrograman dibagi menjadi modul ataupun komponen; berikut adalah simbol pada *component* diagram;

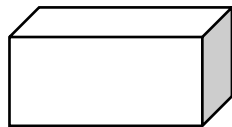


(f) *deployment diagram*;

deployment diagram digunakan untuk mendeskripsikan *node hardware* dan *software* dalam sistem dengan menggambarkan konfigurasi komponen *software real-time*, *processor* dan peralatan yang membentuk arsitektur sistem; berikut adalah simbol pada *deployment diagram*;



; *package* merupakan suatu bungkusan dari satu atau lebih komponen;



; *node* mengacu pada *hardware* dan *software*;



; *dependency* menunjukkan kebergantungan antar komponen, arah panah mengarah pada komponen yang digunakan;



; *link* menunjukkan relasi antar *node*.

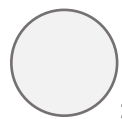
5. Business Process Model and Notation (BPMN)

Dalam menggambarkan alur kerja, diperlukan notasi untuk memudahkan dalam perancangan sistem maupun proses bisnis yang akan dilaksanakan. Menurut (Pant & Juric, 2008, p. 111) *Business Process Model and Notation* (BPMN) merupakan notasi standar yang dirancang untuk memetakan sebagian besar jenis proses yang digunakan dalam suatu organisasi; sebagian besar organisasi menggunakan BPMN untuk membuat suatu proses repositori untuk semua prosesnya. Oleh karena itu, dengan BPMN berbagai proses dipetakan dan dikategorikan sesuai dengan penggunaan bisnis.

Menurut (Pant & Juric, 2008, p. 111), berikut adalah element inti pada BPMN;

(a) *events*;

events menentukan bagaimana suatu organisasi akan merespons situasi melalui proses bisnisnya dengan menunjukkan alur, atau akan memberikan hasil; *events* terbagi dalam 3 tipe, yaitu *start*, *intermediate*, dan *end*; berikut adalah notasi *events*;



(b) *activities*;

activities menunjukkan serangkaian kegiatan yang dilakukan untuk mencapai tujuan yang memiliki titik awal dan titik akhir yang jelas dalam sebuah proses; berikut adalah notasi *activities*;



(c) *sub-process*;

sub-process akan direpresentasikan dengan tanda plus di dalamnya persegi panjang yang menunjukkan bahwa aktivitas yang diberikan dijabarkan lebih lanjut dalam diagram terpisah; berikut adalah notasi *sub-process*;



(d) *gateway*;

gateway digunakan untuk mengontrol bagaimana alur interaksi saat bertemu dan menyimpang di dalam proses; *gateway* dibagi menjadi beberapa jenis, yaitu *exclusive decision/merge(XOR)*, *data-based*, *event-based*, *inclusive decision/merge(OR)*, *complex decision/merge*, *parallel fork/join(AND)*; berikut adalah notasi *gateway*;



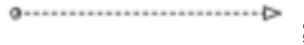
(e) *sequence flow*;

sequence flow digunakan untuk menunjukkan urutan aktivitas yang dilakukan dalam proses; berikut adalah notasi *sequence flow*;



(f) *message flow*;

message flow digunakan untuk menunjukkan aliran pesan yang menghubungkan 2 objek yang dilakukan dalam proses, biasanya *message flow* dapat menghubungkan antar *pool* atau *swimlanes*; berikut adalah notasi *message flow*;



(g) *association*;

association digunakan untuk menghubungkan antara informasi dan objek dengan menampilkan informasi teks tambahan pada suatu aktivitas, teks dan aktivitas dapat dikaitkan menggunakan *association*; berikut adalah notasi *association*;



(h) *pool*;

pool digunakan untuk mengelompokkan berbagai aktivitas yang dilakukan oleh partisipan yang berbeda dalam suatu proses; berikut adalah notasi *pool*;



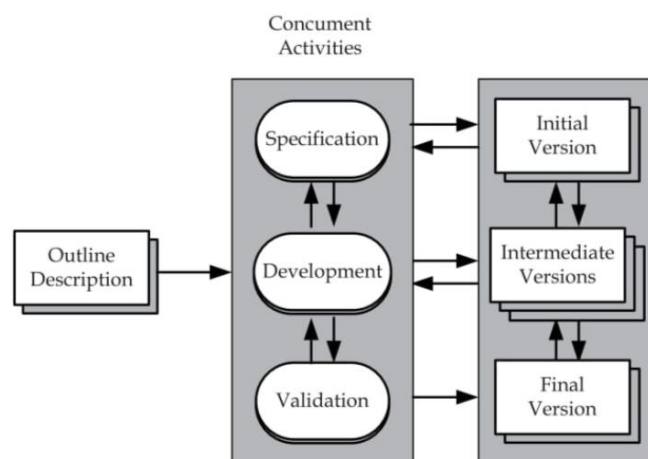
(i) *lanes*;

lanes digunakan untuk menunjukkan kategori lebih lanjut dengan membagi *pool* dalam beberapa bagian; berikut adalah notasi *lanes*;



6. SDLC - *Throwaway Prototyping*

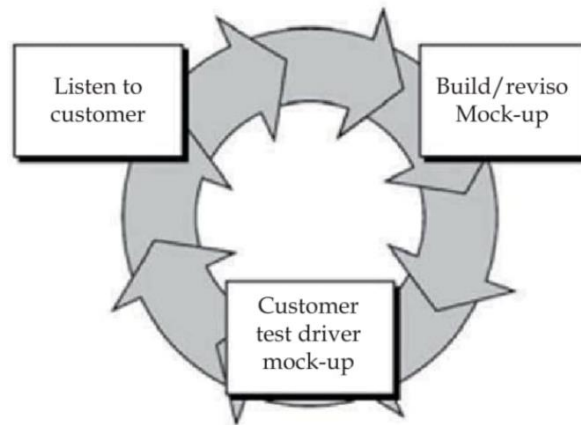
Menurut (Rosa & Salahudin, 2018), *System Development Life Cycle* (SDLC) merupakan penggunaan suatu model dan metodologi untuk melakukan proses pengembangan atau memodifikasi sistem perangkat lunak. Dalam SDLC terdapat beberapa metode, salah satunya adalah metode *prototyping*. Menurut (Kusnadi, Supiandi, Syabaniah, & Oktapiani, 2020, p. 16), *prototyping* merupakan model pengembangan evolusioner berdasarkan gagasan untuk mengembangkan implementasi awal, menunjukkannya kepada pengguna untuk komentar, dan secara bertahap meningkatkan hingga memiliki sistem yang memenuhi persyaratan.; *throwaway prototype* merupakan salah satu jenis model pengembangan *prototype* dengan tujuan untuk memahami apa yang dibutuhkan pelanggan dan menentukan kebutuhan yang lebih baik untuk sistem; berikut adalah model *prototyping*;



Gambar 2. 1 Model Prototyping

Sumber : (Kusnadi, Supiandi, Syabaniah, & Oktapiani, 2020).

Penelitian ini menggunakan *throwaway prototyping* karena *prototype* fokus pada eksperimen tanpa pemahaman menyeluruh tentang kebutuhan pelanggan. Model *prototype* menurut Roger S. Pressman pada (Kusnadi, Supiandi, Syabaniah, & Oktapiani, 2020) terdiri dari tiga tahap, yaitu (1) mendengarkan pelanggan, (2) merancang dan membuat *prototype*, (3) uji coba;



Gambar 2. 2 Tahap Pengembangan Prototype

Sumber : (Kusnadi, Supiandi, Syabaniah, & Oktapiani, 2020);

pada tahap pertama, dikumpulkan kebutuhan pada sistem dengan mengetahui kebutuhan dari pelanggan dengan cara mengetahui bagaimana sistem yang berjalan sebelumnya untuk diketahui masalah yang terjadi sehingga dapat dirancang suatu sistem yang sesuai kebutuhan; selanjutnya, pada tahap kedua dilakukan perancangan *prototype* sistem, dari pendefinisian kebutuhan sistem yang sudah dilakukan sebelumnya, dibuat *prototype* yang sesuai dengan kebutuhan sistem; dan pada tahap terakhir, *prototype* diuji coba oleh pengguna dan dievaluasi kekurangan dari kebutuhan pelanggan sehingga kembali pada tahap mendengarkan keluhan dari pelanggan untuk memperbaiki *prototype*.

7. Bahasa Pemrograman

Menurut (Kurniawan, 2020, p. 30), python adalah *interpreted language*, yang artinya *source code* yang ditulis akan dijalankan per-baris oleh *python interpreter*; python adalah bahasa yang populer digunakan untuk *machine learning* dan *artificial intelligence*; python adalah bahasa *scripting* yang artinya bisa dieksekusi tanpa perlu *compiler*; python memiliki banyak *library* tambahan untuk manipulasi dan visualisasi data. Python digunakan pada penelitian ini karena kemampuannya untuk memanipulasi data dan pengolahan data untuk analisis.

8. Web Server

Menurut (Arief, 2011), *web server* adalah program aplikasi yang memiliki kemampuan untuk menyimpan dokumen web. Semua dokumen web yang ditulis menggunakan skrip sisi klien atau skrip sisi server disimpan di direktori root server web (root dokumen); contoh web server adalah Apache; server web ini biasanya tersedia dalam satu paket dengan server database di dalam aplikasi yang membantu proses pengembangan program..

9. Pemetaan Daerah

Menurut (Spasser, 1997, p. 78), peta adalah alat relasional yang menyediakan informasi antara hubungan entitas yang dipetakan; hal ini memungkinkan peta sains untuk menggambarkan dan memberi makna pada hubungan spasial antara batas penelitian dan area aktivitas yang dapat didistribusikan, termasuk area aktivitas utama; peta sains tidak hanya sebagai alat praktis untuk mengkomunikasikan informasi tentang kegiatan ilmiah, tetapi juga berfungsi sebagai dasar untuk mempelajari atau memahami kegiatan ilmiah dengan menggambarannya secara terstruktur dan terstruktur; visualisasi ilmiah dapat diwujudkan dalam bentuk peta, sehingga memunculkan bidang pemetaan pengetahuan; pemetaan pengetahuan dapat dilakukan dengan beberapa cara yang berkaitan erat dengan topik dokumen; mengacu pada pembahasan ini, pemetaan wilayah adalah kegiatan menggambarkan suatu wilayah berdasarkan aktivitasnya.

10. Kekerasan Terhadap Perempuan dan Anak

Pada “Deklarasi tentang Penghapusan Kekerasan Terhadap Perempuan (1993)”, kekerasan terhadap perempuan merupakan tindakan kekerasan berbasis gender yang mengakibatkan dampak buruk terhadap perempuan baik secara fisik maupun mental; tindakan kekerasan dapat berupa ancaman, pemaksaan maupun perampasan kebebasan perempuan secara serampangan yang dapat terjadi dimana saja baik di tempat umum maupun di kehidupan pribadi.

Kekerasan terhadap anak sesuai dengan Pasal 1 angka 15a Undang-Undang Republik Indonesia Nomor 35 Tahun 2014 tentang Perlindungan Anak menyebutkan bahwa kekerasan terhadap anak adalah setiap perbuatan terhadap anak yang mengakibatkan kesakitan fisik, psikis atau seksual, penderitaan dan/atau penelantaran, termasuk ancaman perilaku, pemaksaan atau perampasan kemerdekaan secara melawan hukum..

B. Tinjauan Pustaka

Dalam pelaksanaan penelitian, dilakukan proses penggalian informasi berdasarkan penelitian yang telah dilakukan sebelumnya. Penelitian mengenai klaterisasi daerah rawan tindak kekerasan terhadap perempuan dan anak sudah pernah dilakukan sebelumnya, begitu pula dengan penelitian yang menggunakan teori Algoritma K-Means. Penelitian tersebut berkontribusi sebagai bahan perbandingan dan sumber pengetahuan penelitian yang dilaksanakan. Berikut adalah penelitian rujukan pada penelitian ini:

(1) Klasterisasi Kasus Kekerasan Terhadap Anak dan Perempuan Berdasarkan Algoritma K-Means (Adawiyah, Sulistiyowati, & Jajuli, 2021)

Penelitian ini menghitung algoritma K-Means menggunakan data kekerasan terhadap anak dan perempuan di Provinsi Kawalan dari tahun 2016 hingga 2020. Menggunakan 3 cluster dalam perhitungan clustering dan evaluasi hasil menggunakan rumus uji kemurnian ukuran menghasilkan nilai kemurnian 0,617 yang menunjukkan cluster sangat baik.

(2) Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Jumlah Perempuan (Usia 18+) Korban Kekerasan Dengan Menggunakan K-Means Clustering (Yulianti, 2021)

Penelitian ini mengelompokkan 35 provinsi/kota di Jawa Tengah menjadi 3 cluster berdasarkan jumlah perempuan (di atas 18 tahun) yang menjadi korban kekerasan. Pengelompokan dilakukan dengan menggunakan metode cluster K-means. Pengelompokan kabupaten/kota di Jawa Tengah menjadi tiga cluster, Klaster 1 – Terdapat satu kabupaten/kota dengan jumlah korban kekerasan perempuan (di atas 18 tahun) yang banyak. Klaster 3 Terdapat 30 kabupaten/kota dengan jumlah korban kekerasan terhadap perempuan yang tinggi (18+) dan jumlah korban kekerasan terhadap perempuan yang rendah (18+)

(3) Sistem Informasi Pemetaan Wilayah Rawan Kriminalitas Polresta Bandar Lampung Menggunakan K-Means Clustering (Nurjoko, Dwirohayanti, & Sudiby, 2020)

Penelitian ini membuat sistem informasi pemetaan wilayah kejahatan berbasis website menggunakan teknik clustering K-Means yang dapat digunakan bersama POLRESTA dan POLSEK untuk mengidentifikasi wilayah kejahatan berdasarkan jenis dan tingkat kejahatan.

(4) Klasterisasi Karakteristik Kekerasan Seksual Terhadap Anak Dengan Metode K-Means Cluster Analysis (Wiza, 2019)

Penelitian ini menganalisis data kasus dari total 335 kasus kekerasan seksual terhadap anak di Pekanbaru. Data yang diperoleh akan sampai dengan 3 cluster.

Klaster yang dominan adalah Klaster 1, dengan kriteria kekerasan seksual terhadap anak usia 3 sampai 16 tahun, dan yang menjadi pemicu terbesar adalah kesempatan.

(5) Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar (Gultom, Tampubolon, Hutabarat, Ilmi, & Hartama, 2020)

Penelitian ini menggunakan algoritma K-Means untuk mengidentifikasi daerah dengan tingkat kriminalitas tinggi dan kriminalitas rendah di wilayah Pematangsiantar. Data yang digunakan dalam penelitian ini berasal dari tahun 2019 dan terdiri dari 6 kecamatan.

(6) Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah (Dewi, Windarto, Damanik, & Satria, 2019)

Penelitian ini menyajikan analisis metode K-Means untuk mengelompokkan kejahatan berdasarkan negara. Sumber data untuk penelitian ini adalah kumpulan dari berbagai dokumen yang menggambarkan aktivitas kriminal di Badan Pusat Statistik. Data yang digunakan dalam penelitian ini adalah data tahun 2012 dari 31 negara bagian. Variabel yang digunakan adalah klaster tingkat tinggi dan klaster tingkat rendah.

(7) Analisis Pemetaan Tingkat Kriminalitas Di Kabupaten Karawang Menggunakan Algoritma K-Means (Fahmi, Jajuli, & Sulistiyowati, 2021)

Penelitian menggunakan teknik clustering menggunakan algoritma K-Means dan pemetaan daerah rawan kejahatan menggunakan QGIS. Hasil evaluasi klasterisasi menggunakan koefisien Silhouette adalah 0,52 pada tahun 2019 dan 0,54 pada tahun 2020, keduanya dalam kategori struktur sedang dengan interpretasi penataan klaster yang wajar.

(8) Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat (Paembonan & Abduh, 2021)

Penelitian ini menggunakan metode K-means, yang digunakan untuk mengurutkan beberapa obat serupa ke dalam kelompok data tertentu. Salah satu cara untuk menentukan kesamaan data adalah dengan menghitung jarak antar data. Tujuan akhir dari clustering adalah untuk menentukan kelompok dalam kumpulan data yang tidak berlabel. Huruf pertama Kriteria untuk jumlah cluster yang dapat dibentuk dalam dataset. , hasil pengelompokan harus dievaluasi. Berdasarkan evaluasi hasil clustered dengan nilai Silhouette Coefficient = 0,4854.

(9) Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia (Tanjung, Windarto, & Fauzan, 2021)

Penelitian ini menggunakan teknik data mining yaitu algoritma K-Means. Metode K-Means adalah teknik pengelompokan yang dirancang untuk menguraikan kumpulan data menjadi kelompok-kelompok. Metode K-Means tersedia untuk data pengangguran negara bagian demi negara bagian. Data dibagi atau dikelompokkan menjadi dua cluster. Di sini, Klaster 1 adalah kelompok negara bagian dengan peluang pengangguran tertinggi dari 13 negara bagian, dan Klaster 2 adalah kelompok negara bagian dengan peluang pengangguran tertinggi. Potensi pengangguran terendah yaitu 21 negara bagian. Hasil penelitian ini diharapkan dapat membantu pemerintah dalam memperluas kesempatan kerja untuk mengembangkan dan meningkatkan perekonomian setiap provinsi di Indonesia.

(10) Data mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia dengan Menggunakan Metode K-Means Clustering (Defit & Nurcahyo, 2021)

Penelitian ini mengukur kemampuan siswa untuk maju melalui proses pembelajaran dan berfungsi sebagai referensi dan sumber evaluasi bagi sekolah dalam hal keberhasilan pendidik dalam melakukan proses belajar mengajar. Dalam penelitian ini, data diperoleh dari SMP Islam Andalas Cendekia Dharmasraya. Ini terdiri dari beberapa variabel: data kehadiran siswa, kinerja akademik (pengetahuan), kinerja psikomotor (keterampilan), dan nilai afektif (mental dan sosial). Untuk pengelompokan data, metode clustering menggunakan algoritma K-Means merupakan metode yang tepat dalam penelitian ini.

Tabel 2. 1 Tinjauan Pustaka

| No | Peneliti | Judul | Jurnal | Kontribusi |
|----|---|--|---|--|
| 1 | Noviya Adawiyah, Nina Sulistiyowati, Mohamad Jajuli (2021) | Klasterisasi Kasus Kekerasan Terhadap Anak dan Perempuan Berdasarkan Algoritma K-Means | Generation Journal /Vol.5 No.2 /e-ISSN: 2549-2233 /p-ISSN: 2580-4952 https://ojs.unpkediri.ac.id/index.php/gj/article/view/15995 | Melakukan klasterisasi daerah rawan tindak kekerasan di kabupaten Karawang menggunakan Weka dengan algoritma K-Means dan metode elbow. |
| 2 | Evi Yulianti (2021) | Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Jumlah Perempuan (Usia 18+) Korban Kekerasan Dengan Menggunakan K- Means <i>Clustering</i> | Jurnal Ilmiah Matematika dan Pendidikan Matematika (JMP) Vol. 13 No. 2, hal. 81-92 ISSN (Cetak) : 2085-1456; ISSN (Online) : 2550-0422 http://jos.unsoed.ac.id/index.php/jmp/article/view/5014/2674 | Menghasilkan 3 <i>cluster</i> daerah berdasarkan jumlah kekerasan terhadap perempuan (Usia 18+) di Jawa Tengah dengan algoritma K-Means |
| 3 | Nurjoko, Defi Dwirohayati, Novi Herawadi Sudiby (2020) | Sistem Informasi Pemetaan Wilayah Rawan Kriminalitas Polresta Bandar Lampung Menggunakan K- Means <i>Clustering</i> | JURNAL TEKNIKA ISSN: 0854-3143 e- ISSN: 2622-3481 https://jurnal.polsri.ac.id/index.php/teknika/article/view/2454 | Menghasilkan sistem informasi pemetaan daerah rawan tindak kriminalitas di Bandar Lampung dengan metode K-Means menggunakan Rapid Miner. |
| 4 | Fana Wiza (2019) | Klasterisasi Karakteristik Kekerasan Seksual Terhadap Anak Dengan Metode K- Means <i>Cluster</i> Analysis | Jurnal Teknologi Informasi & Komunikasi Digital Zone, Volume 10, Nomor 1 , Mei 2019: 44- 53 https://journal.unilak.ac.id/index.php/dz/article/view/2423 | Menghasilkan pola kelompok tindak kekerasan seksual terhadap anak yang paling dominan terjadi disertai kriteria yang mendominasi. |

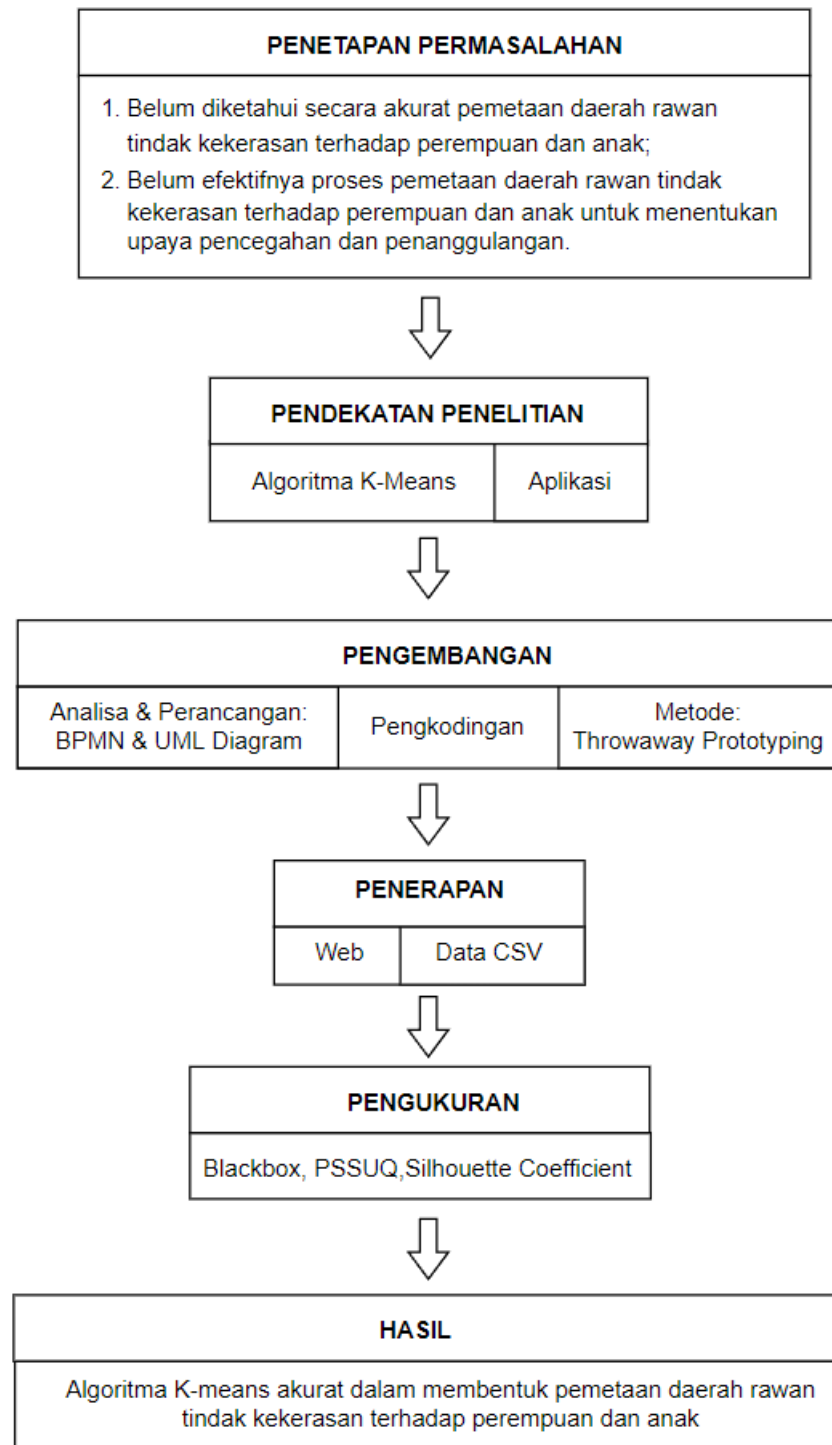
| No | Peneliti | Judul | Jurnal | Kontribusi |
|----|--|---|---|---|
| 5 | Devi Gultom, Hotma Dame Tampubolon, Luvita Yolanda Hutabarat, Fikrul Ilmi R H Zer, Dedy Hartama (2020) | Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar | (Jurnal Teknologi Informasi) Vol.4, No.1, Juni 2020 P-ISSN 2580- 7927 E-ISSN 2615- 2738 http://jurnal.una.ac.id/index.php/jurti/article/view/1263 | Menghasilkan 2 <i>cluster</i> daerah berdasarkan tingkat kejahatan di Pematangsiantar dengan algoritma K- Means |
| 6 | Sinta Maulina Dewi, Agus Perdana Windarto, Irfan Sudahri Damanik, Heru Satria (2019) | Analisa Metode K- Means pada Pengelompokan Kriminalitas Menurut Wilayah | ISBN: 978-602-52720-2- 8 SENSASI 2019 https://seminar-id.com/prosiding/index.php/sensasi/article/view/376/368 | Menghasilkan 2 <i>cluster</i> daerah berdasarkan tingkat kriminalitas menurut provinsi dengan algoritma K- Means |
| 7 | Resti Noor Fahmi, Mohamad Jajuli, Nina Sulistiyowati (2021) | Analisis Pemetaan Tingkat Kriminalitas Di Kabupaten Karawang Menggunakan Algoritma K-Means | Journal of Information Technology and Computer Science (INTECOMS) Volume 4 Nomor 1, Juni 2021 e-ISSN : 2614- 1574 p-ISSN : 2621- 3249 https://journal.ipm2kpe.or.id/index.php/INTECOM/article/view/2413 | Menghasilkan pengelompokan daerah rawan kriminalitas di Kabupaten Karawang menjadi 3 kluster dengan menggunakan algoritma K-Means dan pengujian <i>silhouette coefficient</i> |
| 8 | Solmin Paembonan, Hisma Abduh (2021) | Penerapan Metode <i>Silhouette coefficient</i> Untuk Evaluasi Clustering Obat | Vol. 6, No.2, September (2021) p-ISSN 2502-8952 e-ISSN 2623-2197 http://ojs.unanda.ac.id/index.php/jiit/article/view/659/573 | Menggunakan pengujian dengan <i>silhouette coefficient</i> untuk mengevaluasi hasil <i>cluster</i> obat dengan jumlah <i>cluster</i> = 2 dan menunjukkan bahwa jumlah <i>cluster</i> ideal. |

| No | Peneliti | Judul | Jurnal | Kontribusi |
|----|---|---|---|--|
| 9 | Fadhillah Azmi Tanjung, Agus Perdana Windarto, M Fauzan (2021) | Penerapan Metode K-Means pada Pengelompokan Pengangguran di Indonesia | Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK) Volume 6 Nomor 1 Februari, pp 61-74 ISSN: 2527- 5771/EISSN: 2549-7839 https://tunasbangsa.ac.id/ejurnal/index.php/jurasik/article/view/271 | Menghasilkan 2 <i>cluster</i> daerah berdasarkan tingkat pengangguran berdasarkan provinsinya dengan menggunakan metode K-Means dan pengujian dengan tools <i>rapidminer</i> |
| 10 | Melissa Triandini, Sarjon Defit, Gunadi Widi Nurchahyo (2021) | <i>Data Mining</i> dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia dengan Menggunakan Metode K-Means <i>Clustering</i> | Jurnal Informasi dan Teknologi Vol.3 No.3 Hal:167-173 ISSN: 2714-9730 https://jdt.org/index.php/jdt/article/view/120 | Menghasilkan 3 <i>cluster</i> sesuai tingkat keaktifan siswa dengan menggunakan metode K-Means dan pengujian <i>Davies Bouldin Index</i> |

Berdasarkan tinjauan pustaka yang sudah dipaparkan pada tabel 2.1, didapatkan pengetahuan yang dijadikan rujukan dalam pelaksanaan penelitian ini. Rujukan penelitian berkontribusi dalam memberi pengetahuan mengenai permasalahan terkait kekerasan terhadap perempuan dan anak, proses perhitungan algoritma K-Means. Penelitian yang dilakukan merupakan pengembangan dari penelitian yang ada sebelumnya. Pada penelitian ini, dilakukan pengembangan penelitian dengan menggunakan metode K-Means dan diuji dengan *silhouette coefficient* untuk melakukan pemetaan daerah rawan tindak kekerasan terhadap perempuan dan anak dengan menggunakan variabel nama daerah, psikis, fisik, seksual, eksploitasi, trafficking, penelantaran dan lainnya..

C. Kerangka Pemikiran

Kerangka pemikiran adalah skema untuk merangkai konsep pemikiran pada penelitian. Gambar 2.1 adalah kerangka pemikiran pada penelitian ini.



Gambar 2. 3 Kerangka Pemikiran

Deskripsi kerangka pemikiran pada gambar 2.3:

- (1) Permasalahan pada penelitian adalah yaitu belum diketahui secara akurat pemetaan daerah rawan tindak kekerasan terhadap perempuan dan anak dan belum efektifnya proses pemetaan daerah rawan tindak kekerasan terhadap perempuan dan anak untuk menentukan upaya pencegahan dan penanggulangan;
- (2) Untuk menemukan solusi dari permasalahan, maka dilakukan pendekatan penelitian dengan metode K-Means;
- (3) Pengembangan penelitian dilakukan dengan melakukan analisa dan perancangan menggunakan UML, kemudian melakukan pengkodean dengan menggunakan bahasa pemrograman python dengan metode *throw-away prototyping*;
- (4) Penerapan penelitian dilakukan dengan menggunakan web dan data CSV;
- (5) Untuk menguji penelitian dan pengembangan yang dilakukan, maka dilakukan pengukuran dengan menggunakan *black box*, PSSUQ, dan *silhouette coefficient*;
- (6) Hasil dari penelitian dan pengembangan adalah terbukti bahwa algoritma K-Means akurat dalam membentuk pemetaan daerah rawan tindak kekerasan terhadap perempuan dan anak.

D. Hipotesis Penelitian

Algoritma *K-Means* merupakan metode *data mining* yang digunakan untuk melakukan klusterisasi dengan mengelompokkan objek berdasarkan kesamaan ataupun ketidaksamaan karakteristiknya. Merujuk pada ketercapaian penelitian sebelumnya, metode K-Means mampu membentuk *cluster* yang akurat. Berdasarkan pada pernyataan masalah belum diketahui secara akurat dan belum efektifnya pemetaan daerah rawan tindak kekerasan terhadap perempuan dan anak, maka diperlukannya klusterisasi daerah rawan tindak kekerasan terhadap perempuan dan anak yang lebih optimal untuk memberikan informasi yang tepat pada pihak terkait untuk mengambil tindakan pencegahan dan penanganan pada permasalahan tersebut. Penerapan K-Means diduga akurat dan efektif dalam melakukan pemetaan daerah rawan tindak kekerasan terhadap perempuan dan anak.