**ANALISIS PEMETAAN DATA KRIMINAL DI WILAYAH DILI DENGAN METODE CLUSTERISASI *K-MEANS* BERBASIS WEBGIS**

****

**DISUSUN OLEH :**

**ELISABET MARIA PEREIRA DA COSTA LEITE**

**04317038**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS NAROTAMA**

**SURABAYA**

**2020**

**BAB I**

**PENDAHULUAN**

* 1. **LatarBelakang**

Timor Leste merupakan salah satu negara yang masih muda setelah berhasil melepaskan diri dari Indonesia dan menjadi Negara demokrasi pada tahun 2002. Secara geografis Timor-Leste terletak di Asia Tenggara, dan merupakan Negara baru yang bergabung di kawasan tersebut. Sebagai negara yang masih dalam proses perkembangan, Timor-Leste belum menerapkan dan memaksimalkan teknologi informasi dalam dunia pekerjaan. Contohnya dalam bidang keamanan terutama di kepolisian, data yang berada di kepolisian masih bersifat data mentah dan berupa data manual. Hal ini kurang efektif dalam mempermudah pekerjaan kepolisian untuk melakukan presentasi mengenai tindak kejahatan dan kriminalitas, karena belum diterapkannya teknologi informasi untuk membantu tugas tersebut.

Kriminalitas atau tindak kriminal merupakan masalah umum yang sering terjadi di perkotaan besar contohnya di kota Dili. Dili merupakan ibu kota Timor-Leste dan juga menjadi salah satu kota terbesar di Timor-Leste. Kebutuhan hidup yang cukup tinggi, kepadatan penduduk yang setiap tahun bertambah, dan banyaknya perbedaan pemikiran antara satu kelompok dengan kelompok yang lain menjadikan hal tersebut sebagai beberapa factor terjadinya suatu tindakan kriminalitas.

Berdasarkan uraian di atas, maka dari itu penulis ingin memanfaatkan perkembangan era digital ini dengan membuat sistem informasi pengolah data kriminal berbasis WebGis menggunakan metode *clusterisasi*. Fungsi sistem informasi geografis (SIG) adalah untuk menampilkan data yang sudah diolah menjadi sebuah aplikasi berbasis pemetaan. Dengan dibuatnya sistem ini diharapkan mampu memberikan infornasi tentang tindakan kriminaitas dengan cepat dan akurat, dan mampu mempermudah pihak berwenang dalam menempatkan anggota sesuai kejahatan yang sudah dipetakan.

* 1. **Rumusan Masalah**

Berdasarkan penjelasan latar belakang di atas, penulis dapat merumuskan masalah sebagai berikut :

1. Bagaimana cara untuk memetakan setiap tindakan kriminal ke dalam Sistem informasi geografis (SIG) berdasarkan wilayah, jam, dan jenis tindakan kriminalitas?
2. Bagaimana penerapan metode *clustering K-means* dalam pemrosesan data kriminal?
   1. **Batasan Masalah**

Untuk menghindari pembahasan terlalu luas dalam penelitian ini maka dilakukan pembatasan terhadap permasalahan yang diambil. Adapun beberapa batasan masalaha dalah sebagai berikut :

1. Sistem pengolah data kriminal hanya terbatas untuk wilayah di kota Dili.
2. Metode yang digunakan dalam penelitian ini adalah *Clustering K-Means.*
3. Terdapat beberapa wilayah di kota Dili yang belum memiliki *longitude* dan *latitude* yang tepat sehingga sulit untuk menentukan titik koordinat di wilayah tersebut.
4. Aplikasi yang digunakan untuk melakukan pemrosesan data adalah *jupyterlab*
5. Bahasa pemrograman yang digunakan adalah python, *output* atau hasil yang akan ditampilkan berupa *choropletmap*.
   1. **Tujuan Penilitian**

Tujuan dari penelitian ini adalah sebagai berikut :

1. Memetakan setiap wilayah berdasarkan data yang sudah di proses menggunakan metode clusterisasi *K-means* dan choroplet map, dengan ini diharapkan bisa membantu pihak kepolisian untuk bisa memberikan informasi kepada masyarakat tentang daerah yang rawan akan tindak kriminal.
2. Untuk mengetahui keakuratan metode clusterisasi *K-means* dalam melakukan pemrosesan data kriminal pada penelitian ini.
   1. **Manfaat Penelitian**

Manfaat yang dicapai dari penelitian ini adalah membantu masyarakat untuk menghindari wilayah yang berpotensi terjadinya tindak kejahatan. Hasil dari penilitian ini juga dapat dijadikan sebagai sarana bagi pihak berwenang untuk mengurangi tindak kejahatan di kota Dili dengan menempatkan anggota kepolisian sesuai dengan tingkat kejahatan di suatu wilayah.

* 1. **Sistematika Penulisan**

Sistematika penulisan dalam penelitian ini terdari dari 5 bab, dimana tiap bab membahas tentang masing-masing ruang lingkup :

**BAB I PENDAHULUAN**

Bab ini menjelaskan tentang latar belakang mengapa penulis mengambil tema atau judul ini, selain itu bab ini juga mencakup sub bab latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

**BAB II TINJAUAN PUSTAKA**

Pada bab ini membahas mengenai penelitian terdahulu yang berkaitan dengan topik penelitian yang sedang dikerjakan dan menguraikan teori-teori dasar pembahasan yang dilakukan dalam penelitian ini.

**BAB III METODOLOGI PENELITIAN**

Pada bab ini membahas tentang alur proses yang ada pada penelitian, yaitu pengolahan data, analisis data, dan perancangan sistem informasi pengolah data kriminal berbasis WebGis.

**BAB IV HASIL DAN PEMBAHASAN**

Bab ini akan memaparkan atau menyajikan hasil dari sistem informasi pengolah data kriminal berbasis WebGis berupa keakuratan data yang sudah diproses menggunakan metode clusterisasi *K-Means* kemudian akan ditampilkan melalui grafik dan map.

**BAB V KESIMPULAN**

Bab ini berisikesimpulan dan saran dari peneliti secara keseluruhan, sehingga dapat menjadi acuan pada penelitian berikutnya.

**BAB II**

**TINJAUAN PUSTAKA**

* 1. **Penelitian Terdahulu**

Tinjauan penelitian pertama disusun oleh Gilang YudistiraHilman, Bandi Sasmito, dan Arwan Putra Wijaya yang berjudul “PEMETAAN DAERAH RAWAN KRIMINALITAS DI WILAYAH HUKUM POLTABES SEMARANG PADA TAHUN 2013 DENGAN MENGGUNAKAN METODE CLUSTERING”. Pada penelitian ini dijelaskan tentang pemetaan daerah rawan kriminalitas berdasarkan kepadatan penduduk dan angka kemiskinan yang tinggi di kota Semarang. Variabel yang digunakan untuk analisa data pada penelitian ini yaitu kepadatan penduduk, presentase kinerja polsek dan juga angka kemiskinan sebagai *independent variable* (variabel X) dan jumlah tindak kejahatan sebagai *dependent variable*(Variabel Y), kemudian hasil analisis dilkukan uji statistk menggunakan uji F. Ada2 hasil yang diperoleh dari penelitian ini yaitu, hasil pertama menunjukan bahwa beberapa wilayah di kota Semarang memilki tingkat kerawanan yang berbebeda, dan hasil kedua yang diperoleh adalah tindak kejahatan paling sering terjadi antara pukul 00.01 - 06.00 kemudian 06.01-12.00, 12.01-18.00 dan 06.00-12.00 di mana 50,03 % tindak kejahatan terjadi antara pukul 00.01-06.00.

Tinjauan penelitian kedua yang disusun oleh Jajang Jaya Purnama, Ridan Nurfalah, dan Hafifah Bella Novitasari dengan judul “ANALISA ALGORITMA K-MEANS CLUSTERING PEMETAAN JUMLAH TINDAK PIDANA”.Penelitian ini menjelaskan tentang penerapan model data mining dalam mengelompokan*illegal fishing* dengan mengimplementasikan algoritma data mining *k-means*. Data yang digunakan dalam penelitian ini merupakan data sampel yang diambil dari dataset global tindak pidana *illegal fishing* dari tahun 2003-2013 dan akan diuji 266 data tindak pidana *illegal fishing.* Kemudian diperoleh kesimpulan bahwa analisis menggunakan algortitma data mining *k-means* dengan penentuan jarak terdekat menggunakan *Euclidean distance* lebih optimal dibandingkan dengan *mahattan distance* dan *chbchep distance*, dan untuk penentuan centroid (titik pusat) pada tahap awal algoritma k-means sangat berpengaruh pada hasil cluster.

Tinjauan penelitian ketiga disusun oleh Ferdian Dwi Yuliansyah dengan judul “SISTEM INFORMASI GEOGRAFIS UNTUK KLASIFIKASI DAERAH RAWAN KRIMINALITAS MENGGUNAKAN METODE K-MEANS ”. Pada penelitian ini menjelaskan tentang pemanfaatan sistem informasi goegrafis dalam pengelompokan daerah-dareah yang rawan terjadinya kriminalitas menggunakan *k-means* dengan parameter jumlah kejadian kriminalitas untuk 5 kategori, yaitu Pencurian, Perjudian, Pemerkosaan, dan penipuan. Terdapat bebebrapa hasil yang diperoleh dari penelitian ini, hasil penelitian pertama adalah Aplikasi Sistem Informasi Geografis menggunakan metode K-means dapat memetakan daerah rawan kriminalitas di Kabupaten Cilacap, hasil penelitian kedua User atau pengguna dapat mengetahui informasi lokasi daerah rawan kriminalitas pasa sistem tersebut dalam bentuk peta, dan hasil yang terakhir adalah masyarakat dapat memantau perkembangan dari kerawanan tindak kriminalitas pada masing-masing daerah berdasarkan informasi dan analisis yangterdapat pada sistem.

Tinjauan penelitian keempat disusun oleh Yulia Darmi, dan Agus Setiawan yang berjudul “PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK” .Dalam penelitian ini menjeleskan tentang pengklasifikasian antara barang yang laku dantidak laku dengan metode *clustering k-means*,hasil yang didapatkan dengan menggunakan metode tersebut adalah mengetahui dengan baik barang laku maupun yang tidak laku sehingga dapat dikurangi pemesanan untuk barang yang tidak laku.

**Tabel 2.1 Penelitian Terdahulu**

|  |  |  |  |
| --- | --- | --- | --- |
| No | Peneliti | Judul | Hasil |
| 1 | Gilang YudistiraHilman, Bandi Sasmito, Arwan Putra Wijaya | PEMETAAN DAERAH RAWAN KRIMINALITAS DI WILAYAH HUKUM POLTABES SEMARANG PADA TAHUN 2013 DENGAN MENGGUNAKAN METODE CLUSTERING | * Beberapa daerah di Kota Semarang memiliki tingkat kerawanan terhadap kejahatan yang berbeda. * Tindak kejahatan paling sering |

**Tabel 2.1 Penelitian Terdahulu Lanjutan**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | terjadi antara pukul 00.01 - 06.00 kemudian 06.01-12.00, 12.01-18.00 dan 06.00-12.00 di mana 50,03 % tindak kejahatan terjadi antara pukul 00.01-06.00. |
| 2 | Jajang Jaya Purnama, Ridan Nurfalah, dan Hafifah Bella Novitasari | ANALISA ALGORITMA K-MEANS CLUSTERING PEMETAAN JUMLAH TINDAK PIDANA | * analisis menggunakan algortitma data mining *k-means* dengan penentuan jarak terdekat menggunakan *Euclidean distance* lebih optimal dibandingkan dengan *mahattan distance* dan *chbchep distance* |

**Tabel 2.1 Penelitian Terdahulu Lanjutan**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | * Penentuan centroid dengan *mahattan distance* dan *chbchep distance*(titik pusat) pada tahap awal algoritma k-means sangat berpengaruh pada hasil cluster. |
| 3 | Ferdian Dwi Yuliansyah | SISTEM IFNORMASI GEOGRAFIS UNTUK KLASIFIKASI DAERAH RAWAN KRIMINALITAS MENGGUNAKAN METODE K-MEANS | * Aplikasi sistem Informasi Geografis menggunakan metode *K-means* dapat memetakan daerah rawan kriminalitas di Kabupaten Cilacap. * Pengguna dapat |

**Tabel 2.1 Penelitian Terdahulu Lanjutan**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | mengetahui informasi lokasi daerah rawan kriminalitas pada sistem tersebut dalam pentuk peta.   * Memantau perkembangan dari kerawanan tindak kriminalitas pada masing-masing daerah dengan adanya informasi tingkat kerawanan kriminalitas dan analisis yang terdapat pada sistem ini. |
| 4 | Yulia Darmi, Agus Setiawan | PENERAPAN METODE CLUSTERING K- MEANS DALAM PENGELOMPOKAN | * Mengetahui dengan baik barang yang laku maupun yang tidak |

**Tabel 2.1 Penelitian Terdahulu Lanjutan**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | laku sehingga pemesanan untuk barang yang tidak laku. |

* 1. **Landasan Teori**
     1. **Data Mining**

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machinelearning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai *database* besar.[1]

Data Mining adalah proses mencari atau menambang data dengan maksud untuk mengenali pola data sehingga dapat menjadi sebuah informasi yang berharga. Data Mining juga termasuk dalam ilmu *Knowladge, Discovery in Database* (KDD).[2]

Berdasarkan definisi-definisi sebelumnya tentang data miningdapat dikatakan bahwa data mining merupakan proses pengamatan suatu data dalam menemukan atau mengenali pola dari data tersebut, sehingga dapat menghasilkan sebuah informasi yang bermanfaat.

* + - 1. **Fungsi Data Mining**

Pada dasarnya data mining memiliki beberapa fungsi yang berperan penting antara lain sebagai berikut :

1. *Clustering*

*Clustering* atau klasterisasi berfungsi untuk membagi data ke dalam sebuah kelompok berdasarkan data yang memiliki tingkat similiaritas atau kemiripan karakter data anyata satu dengan yang lain.

1. *Classification*

*Classification* merupakan salah satu fungsi dari data mining untuk menentukan *record* data baru yang akan dijadikan sebagai *label* atau *class*.

1. *Association*

Assosiasi merupakan menemukan pola atribut dari sebuah record data transaksi yang muncul dalam suatu waktu.Dalam dunia bisnis dikenal sebagai *(Market Basket Analysis*) guna dapat mengetahui suatu keranjang belanja *(behavior) customer* berdasarkan item yang dibeli.[3]

1. *Predicting*

*Predicting* merupakan proses untuk memprediksi kejadian yang terjadi di masa akan datang berdasarkan data-data yang sudah diperoleh dan diolah dengan menggunakan beberapa algoritma seperti apriori maupun *fp-growth.*

* + 1. ***Clustering***

Pada dasarnya *clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan*(similarity)* karakterisktik antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode data *mining* yang bersifat tanpa arahan *(unsupervised),* maksudnya metode ini diterapkan tanpa adanya latihan *(taining)* dan tanpa ada guru *(teacher)* serta tidak memerlukan target output. Dalam data *mining* ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *nonhierarchical clustering.*[4]

*Clustering* atau analisis pengelompokkanmerupakan suatu proses membagi data dari suatu himpunan tertentu kedalam beberapa kelompok yang mempunyai kesamaan karakteristik data dengan karakteristik data dalam kelompok lain. Kelas yang terdapat dalam proses *clustering* dinamakan dengan cluster. Cluster merupakan suatu kelompok data yang mempunyai tingkat kesamaan dengan suatu cluster dan karakteristiknya tidak sama dengan cluster lainnya.[5]

Dengan demikian dapat dikatakan bahwa *clustering* adalah metode dari data minig yang bersifat tidak teratur yang bertujuan untuk mengelompokan data berdasarkan nilai dari atribut ke dalam suatu kelompok.

* + 1. **K-Means**

*K-Means* merupakan salah satu algoritma dari analisis cluster non hirarki yang bertujuan untuk mempartisi data kedalam beberapa kelompokclusterberdasarkan karakteristiknya, sehingga data yang memiliki karakteristiksama akan digabungkan ke dalam satu cluster.

*K-Means* merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*,yang pada umumnya berusaha meminimalisasikan variasi di dalam suatucluster dan memaksimalisasikan variasi antar cluster.[6]

Pada dasarnya analisis cluster menggunakan metode *K-means* memiliki algoritma sebagai berikut : [7]

1. Menentukan banyanknya k cluster yang diinginkan.
2. Menentukan pusat cluster *(centroid)* awal secara *random* dari objek data sebanyak k.
3. Menentukan jarak terdekat dari setiap objek pengamatan dengan *centroid* yang telah ditentukan menggunakan jarak euclidean dengan persamaan (2.1)

(2.1)

Keterangan :

*d* = jarak

*i* = banyaknya data

*c* = centroid

*x* = data

1. Menentukan jarak terdekat suatu objek dengan *centroid.*
2. Menentukan pusat *cluster (centroid)* baru dengan menghitung rata-rata masing-masing *cluster* menggunakan persamaan (2.2).

……………. (2.2)

Keterangan:

𝑐𝑘𝑙 : Nilai *centroid* ke-k pada variabel ke-l.

𝑝 : Banyaknya data

1. Jika suatu objek tidak berpindah cluster dari jarak *centroid* lama dan jarak *centroid* baru maka ulangi langkah 3.
   * 1. **Sistem Informasi Geografis**

Sitem informasi geografis adalah sistem yang berfungsi untuk mengolah, menganalisa maupun menampilkan data yang bersifat spasial yaitu sebuah data yang berorientasi geografis. Teknologi yang dimilik oleh SIG dapat mengoperasikan *database* seperti analisa statistk dan *query,* dengan keunikan visualisasi dan analisa yang dimiliki.

Sistem Informasi Geografis (SIG) adalah suatu sistem yang berfungsi untuk data yang terefrensi secara spasial.SIG juga merupakan sistem informasi yang digunakan untuk menampilkan informasi secara grafis dengan bantuan peta sebagai dasarnya.Dalam SIG terdapat lapisan *(layer)* dan hubungan yang terkonsep. SIG mempunyai kemampuan dasar untuk menganalisis data contohnya query, menganalisis juga menampilkan dalam bentuk pengelompokkan berdasrkan letak geografisnya.[8]

Sistem Informasi Geografis (SIG) mempunyai 4 subsistem yaitu adalah sebagai berikut :[8]

1. Data Masukan

Data masukan berfungsi untuk mengumpulkan, mempersiapkan data spasial dan atribut dari berbagai serta memiliki tujuan untuk mengubah format data awal menjadi format yang dipakai SIG.

1. Data Keluaran

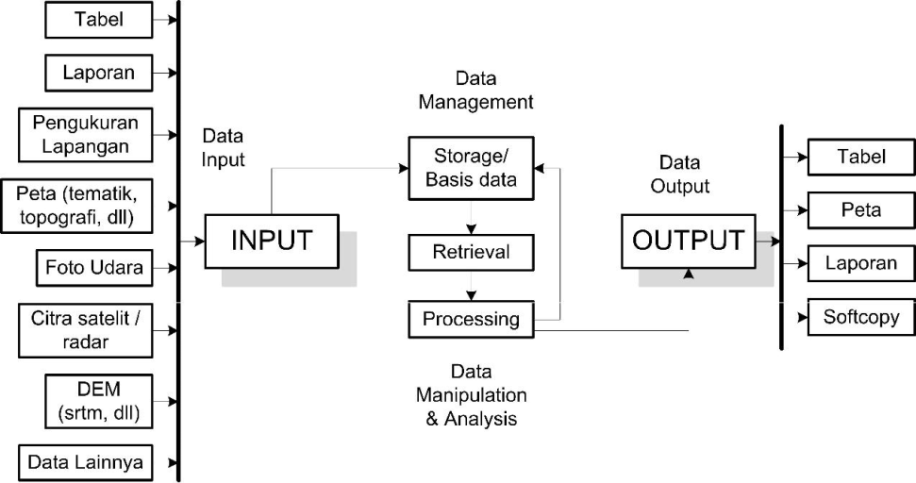
Data Keluaran berfungsi untuk menampilkan hasil pengeluaran secara menyeluruh maupun sebagian basis data baik berupa softcopy atau hardcopy seperti grafik, peta, tabel, dan lain-lain.

1. Data Manajemen

Data Managemen berfungsi untuk mengorganisasikan data spasial atau atribut ke dalam sebuah basis data sehingga mudah diperbarui, diupdate dan dipakai.

1. Data Manipulasi dan Analisis

Sebsistem ini berfungsi untuk memodelkan dan memanipulasi data sehingga menjadi informasi yang diingankan dan menentukan beragai informasi yang dihasilkan oleh SIG. Berikut adalah Gambar 2.1 Ilustrasi Sub-sistem SIG



**Gambar 2.1** Ilustrasi Sub-Sistem SIG

* + 1. **Python**

Python diciptakan oleh Guido van Rossum pada tahun 1990-an. Python merupakan salah satu bahsa pemrograman yang berorientasi Objek *(object oriented programming)* dan banyak digunakan untuk pengembangan *software* maupun untuk analisa data oleh programmer dan developer karena python merupakan bahasa pemrograman yang bersifat *open source*, dan mudah dipahami. Pada analisa data *(Data Analyst)* python berperanpenting karena python memiliki banyak *library* yang bisa dimanfaakan dalammelakukan load data, membuat grafik, danmelakukan proses lain yang dibutuhkan dalam melakukan analisis data.[9]

* + 1. **JupiterLab**

Jupiterlab merupakan salah satu *tool* yang digunakan untuk mengolah data dengan menggunakan bahasa pemrograman python.[10]. Tool ini memeliki beberapa fitur yang sangat membantu seseorang untuk mempermudah dalam melakukan pengolahan data, contohnya mengimport file dalam bentuk csv maupun excel kemudian menampilkan secara rapi, jupyterlab dapat digunakan untuk membuat berbagai dokumen yang berisi kode secara interaktif seperti dapat memvisualisasikan hasil dari data yang sudah diolah kedalam bentuk grafik ataupun chart sehingga bisa memudahkan pengguna untuk mengetahui maupun menganalisa hasilyang diperoleh dari pengolahan data tersebut.

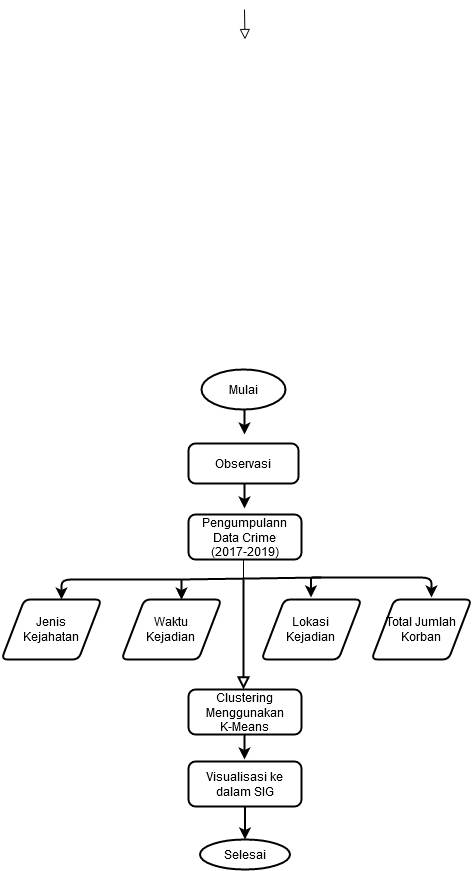
* + 1. **Flask**

Flask merupakan kerangka web *(web framework)* yang ditulis dengan bahasa pemrograman python dan juga termasuk jenis *micro framewok*[11] yang bertujuan untuk menjaga inti aplikasi agar tetap sederhana namun dapat diperluas. Flask *web framework*menyediakan pustaka atau *library* yang siap digunakan untuk membangun sebuah aplikasi web. Pada tahun 2010 Armin Ronacher mulai mengembangkan flask dan mulai bisa digunakan secara luas untuk membuat berbagai sistem maupun *software*.Salah satu aplikasi web yang dibuat dengan menggunakan flask adalah linkedln, pinterest.

**BAB III**

**METODOLOGI PENELITIAN**

Metode penelitian merupakan tahapan yang menggambarkan proses dalam sebuah penelitian yang sedang dilakukan agar penelitian tersebut dapat dilakukan secarat terstruktur sehingga bisa mencapai hasil yang diharapkan. Berikut adalah gambar 3.1 diagram *flowchart* desain penelitian.



Gambar 3.1 Diagram flowchart desain penelitian

## Observasi

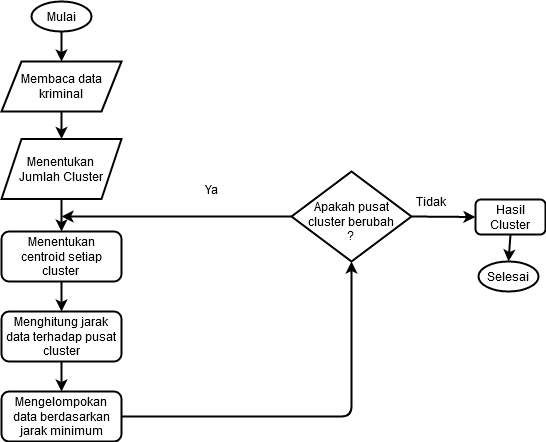
Pada tahapan penelitian ini peneliti melakukan observasi pada kepolisian pusat yang berada di Dili, Timor-Leste. Observasi yang dilakukan yaitu mengenai sistem dan juga data kriminal yang ada di wiliyah kepolisian tersebut, dengan dilakukannya observasi ini peneliti bisa mengetahui secara langsung dalam pengumpulan data kriminal dan juga sistem apa yang akan dibuat pada penelitian kali ini. Sehingga hal ini memudahkan bagi peneliti untuk membuat sistem yang mudah digunakan dan mudah dipahami oleh *users* atau pengguna.

## Pengumpulan Data

Tahapan ini berfungsi untuk memperoleh informasi dari perusahaan atau instansi yang bersangkutan sebagai bahan analisis pada penelitian ini. Pengumpulan data yang dilakukan pada penelitian yaitu mengumpulkan data kejahatan dari tahun 2017-2019 yang meliputi data jenis kejahatan, waktu kejadian, lokasi kejadian, dan total jumlah korban kejahatan di kota Dili, Timor-Leste. Pada penelitian ini data yang digunakan merupakan data sekunder, Data sekunder adalah kumpulan data yang diperoleh secara tidak langsung, atau dalam hal ini peneliti tidak langsung melakukan pengambilan data dari lokasi kejadian secara langsung melainkan data yang diperoleh adalah hasil dari berbagai instansi yang berkaitan, contoh pada penelitian ini menggunakan data kriminal perkasus tahun 2017-2019 dari kepolisian Dili, Timor-Leste.

* 1. **Clustering dengan *K-means***

Clustering K-means merupakan sebuah metode atau proses yang tujuannya adalah untuk mengelompokan data dengan tingkat kemiripan sama ke dalam satu atau lebih kelompok. Data yang digunakan dalam penelitian ini adalah data kriminalitas yang diambil dari tahun 2017-2019. Untuk melakukan pengolahan pada data tersebut menggunakan *clustering k*-*means* terdapat beberapa langkah yang harus dilakukan, pada gambar 3.2 berikut akan digambarkan *flowchart* dari langkah-langkah yang akan dilakukan pada *clustering k-means*.



Gambar 3.2 Flowchart *clustering K-means*

* + 1. **Clustering K-Means**

Pada gambar 3.1 bisa dijabarkan bahwa proses yang digunakan dalam metode *clustering k-means* pada sub bab ini adalah sebagai berikut.

1. Menentukan banyaknya jumlah K *cluster* yang akan digunakan untuk melakukan pengelompkkan data. Banyaknya *cluster* harus lebih dari banyaknya data (k < n).
2. Menentukan nilai awal *centroid*, terdapat banyak metode yang digunakan untuk menentukan nilai awal *centroid*, namun pada proses kali ini metode yang dipakai adalah mengambil dari sumber data secara *random* atau acak. Adapun metode yang digunakan dalam menentukan nilai awal cluster yaitu adalah metode Elbow, dimana metode tersebut berfungsi untuk mengetahui atau menentukan total jumlah cluster yang tepat untuk digunakan dalam pengujian pada peneltian ini berdasarkan SSE *(Sum Square Error).* Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalammenentukan jumlah cluster terbaik dengan caramelihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik.[12]
3. Menghitung jarak dari setiap data pada pusat *cluster*, sehingga ditemukan nilai atau jarak yang terdekat dengan pusat *cluster*. Rumus-rumus yang digunakan untuk menghitung jarak antara lain sebagai berikut :
4. Euclidean Distance
5. Manhattan
6. Minkowski

Rumus yang digunakan untuk menghitung ada proses kali ini adalah rumus *EuclideanDistance* :

…………… (3.1)

Keterangan :

*d*= jarak

*i* = banyaknya data

*c* = data centroid

*x* = data record

1. Setelah menghitung jarak data, selanjutnya adalah mengelompokan data sesuai dengan clusternya yaitu mengambil data yang memiliki jarak terkecil dari pusat *cluster*.
2. Melakukan pengecekan terhadap nilai pusat cluster apakah terjadi perubahan dari nilai tersebut, jika ada maka akan mengulang kembali ke proses nomor 2 yaitu menentukan nilai awal centorid, jika tidak maka proses akan selesai.
   1. **Visualisasi Sistem Informasi Geografis**

Visualisasi yang dilakaukan pada tahapan ini yaitu, adalah proses untuk menampilkan Sistem informasi geografis dan jenis tindakan kriminalitas yang ada di wilayah Dili, Timor-Leste, dengan merepresentasikan hasil dari data yang sudah diolah pada proses sebelumnya kedalam WebGis berupa lokasi atau peta dan juga grafik dari tindakan kriminal itu sendiri sehingga bisa memudahkan pihak kepolisian atau pihak berwenang untuk membaca informasi. Berikut adalah tahapan yang dilakukan dalam proses visualisasi WebGis.

1. Menginstall *tools* maupun *library* yang dibutuhkan pada penelitian ini.
2. Memasukan data kriminal per tahun 2017-2019, yang sudah diolah pada proses *clustering k-means*.
3. Merepresentasikan kedalam Sistem Informasi Geografis yaitu berupa maps wilayah Dili, map atau peta yang akan ditampilkan adalah peta choroplet, dikarenakan penggunaan peta choroplet ini cocok digunakan untuk melakukan pemetaan wilayah, selain pemetaan visualisasi pada penelitian ini juga direpresentasikan dalam bentuk grafik dari jenis kejahatan, lokasi kejadian, maupun waktu kejadian.

**DAFTAR PUSTAKA**

[1] E. Turban, J. Aronson, and T. Llang, *Decision Support Systems and Intelligent Systems*. 2003.

[2] B. Deekshatulu and P. Chandra, “Classification of heart disease using artificial neural network and feature subset selection,” *Glob. J. Comput. Sci. Technol.*, 2013.

[3] K. Deteksi *et al.*, “Klasifikasi deteksi dini penyakit jantung koroner menggunakan modifikasi bobot hidden layer pada metode extreme learning machine dengan gain ratio,” 2020.

[4] S. Budi, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. 2007.

[5] F. Febrianti, M. Hafiyusholeh, and A. H. Asyhar, “PERBANDINGAN PENGKLUSTERAN DATA IRIS MENGGUNAKAN METODE K-MEANS DAN FUZZY C-MEANS,” *J. Mat. “MANTIK,”* 2016, doi: 10.15642/mantik.2016.2.1.7-13.

[6] Yudi Agusta, “K-Means – Penerapan, Permasalahan dan Metode Terkait,” *J. Sist. dan Inform.*, 2007.

[7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.

[8] E. Prahasta, *Sistem Informasi Geografis: Konsep-konsep Dasar (Perspektif Geodesi dan Geomatika) Edisi Revisi*. 2014.

[9] C. M. Lewandowski, N. Co-investigator, and C. M. Lewandowski, *Python for data Analysis*. 2015.

[10] wikipedia, “Anaconda (Python distribution),” *wikipedia*. 2019.

[11] A. Ronacher, “Flask (A Python Microframework),” *http://flask.pocoo.org/*, 2013. .

[12] T. S. Madhulatha, “AN OVERVIEW ON CLUSTERING METHODS,” *IOSR J. Eng.*, 2012, doi: 10.9790/3021-0204719725.