**BAB I**

**PENDAHULUAN**

# **LATAR BE LAKANG**

Penyakit merupakan suatu hal sering dialami bagi semua manusia mulai dari bayi, balita, remaja hingga lansia, maka setiap manusia akan sulit dalam melaksanakan aktivitas sehari-hari mulai dari penyakit ringan hingga berat. Pada umumnya penyakit ditimbulkan dari cuaca buruk, alergi bahkan keturunan dari keluarga namun juga ada faktor lain yang menunjang timbulnya penyakit yaitu pola hidup tidak sehat serta lingkungan tempat tinggal.

Puskesmas donomulyo adalah salah satu unit kesehatan masyarakat dikecamatan bumi agung. Dari puskesmas donomulyo terdapat data berupa jumlah penyakit yang ada di daerah tersebut, namun banyak nya data pasien tersebut berbanding lurus dengan sulitnya mencari penyakit yang paling dominan. Oleh karena itu dilakukannya penelitian guna mengetahui jumlah penyakit yang sering diderita oleh masyarakat dengan menggunakan data mining dengan algoritma K-Means yang mebantu dalam mengelompokka penyakit yang paling dominan dari data yang ada di puskesmas donomulyo.

Pengelompokkan (*Clustering*) penyakit tersebut berdasarkan jenis penyakit yang paling banyak didalam data. Pengelompokkan penyakit menggunakan Algoritma *K-Means* guna mengetahui jenis penyakit terbanyak yang diderita masyarakat sekitar puskesmas, dan mengetahui persamaan pola data dalam pengolahan menggunakan *software WEKA.* Oleh karena itu dilakukan penelitian dengan judul “PENERAPAN ALGORITMA K-MEANS UNTUK JUMLAH PENYAKIT YANG SERING DIDERITA PADA PUSKESMAS DONOMULYO”**.**

Penelitian ini diharapkan dapat membantu pihak puskesmas untuk mengetahui penyakit yang paling sering diderita sehingga hasil dari penelitian ini dapat digunakan dalam pertimbangan untuk menentukan pola hidup yang sehat dalam masyarakat sehingga dapat mengurangi kemungkinan terjangkit penyakit tersebut dan dapat menurunkan angka penderitanya.

* 1. **RUMUSAN MASALAH**

Adapun rumusan masalah dalam penelitian ini adalah

1. Bagaimana menerapkan Algoritma K-Means pada data puskesmas dalam mengidentifikasikan penyakit yang paling dominan
2. Bagaimana menemukan dan mengelompokkan penyakit berdasarkan data dari puskesmas.
3. Bagaimana menentukan pola hidup sehat berdasarkan hasil dari penelitian.
   1. **BATASAN MASALAH**

Berdasarkan latar belakang masalah diatas terdapat batasan masalah berupa:

1. Penelitian ini hanya pada Puskesmas Donomulyo
2. Hanya sebatas mengelompokkan penyakit yang dominan pada Puskesmas Donomulyo
3. Penelitian ini menggunakan Algoritma K-Means
   1. **TUJUAN**
4. Menerapkan Algoritma K-Means pada data pasien Puskesmas Donomulyo
5. Memahami dan mengetahui pola penyakit yang dominan
6. Dapat mengelompokkan jumlah penyakit yang paling dominan
7. Diharap dapat digunakan dalam pertimbangan menentukan pola hidup sehat

**BAB II**

**LANDASAN TEORI**

1. **Tinjauan Pustaka**

Dalam penelitian ini memerlukan tinjauan pustaka yang mana nantinya dapat membantu dalam penelitian ini, berikut merupakan tinjauan pustaka yang di gunakan dalam penelitian :

|  |  |  |  |
| --- | --- | --- | --- |
| No. Literatur | Nama Peneliti | Tahun | Judul |
|  | Palasian D P Silitongadan Romanus damanik | 2016 | IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA ANALISIS PENYEBARAN PENYAKIT PASIEN PENGGUNA BADAN PENYELENGGARA JAMINAN SOSIAL KESEHATAN (BPJS) |
|  | Hendro Prasetyo, Kuntoro, Windhu Purnomo, Soenarnatalina,Merry Adriani, Bambang Wijanarko | 2014 | Penerapan Clustering Bootstrap dengan Metode K-Means |
|  | Joanna Ardhyanti Mita Nugraha, Yupie Kusumawati | 2016 | *DATA MINING* DENGAN METODE *CLUSTERING* UNTUK PENGOLAHAN INFORMASI PERSEDIAAN OBAT PADA PUSKESMAS PANDANARAN SEMARANG |
|  | Taslimdan Fajrizal | 2016 | Penerapan *algorithma k-mean* untuk *clustering* data obat pada puskesmas rumbai |
|  | Aline Embun Pramadhani dan Tedy Setiadi | 2014 | PENERAPAN DATA MINING UNTUK KLASIFIKASI PREDIKSI PENYAKIT ISPA (*Infeksi Saluran Pernapasan Akut*) DENGAN ALGORITMA DECISION TREE (ID3) |
|  | Randi Rian Putra dan Cendra Wadisman | 2014 | IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA *K-MEANS* |

1. **Palasian D P Silitonga dan Romanus damanik**

Dengan diberlakukannya program BPJS Kesehatan dapat dipastikan bahwa jumlah pasien RSUP Haji Adam Malik Medan semakin meningkat. Seiring dengan meningkatnya jumlah pasien rumah sakit, diharapkan pihak rumah sakit dapat mengetahui klasifikasi penyakit yang ada di masyarakat serta penyebab penyakit tersebut. Dengan demikian pihak rumah sakit khususnya RSUP Haji Adam Malik Medan dapat memberikan masukan bagi pemerintah propinsi Sumatera Utara berkaitan dengan usaha pencegahan penyakit dan penyuluhan kesehatan ke daerah-daerah. Salah satu metode yang dapat dilakukan untuk mengelompokkan data pada basis data adalah metode *K-Means Clustering*.

Metode *K-Means Clustering* membagi data menjadi beberapa kelompok serta dapat menerima masukan berupa data tanpa label kelas . 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 yang berbeda dikelompokkan ke dalam kelompok yang lain.

Hasil dari penelitian ini diharapkan dapat memberikan masukan bagi pemerintah propinsi Sumatera Utara berkaitan dengan usaha pencegahan penyakit dan penyuluhan kesehatan ke daerah-daerah sedangkan bagi pihak RSUP Haji Adam Maslik Medan dapat digunakan sebagai masukan bagi proses antisipasi prioritas pelayanan jika diketahui pola penyakit dengan kencederungan tertinggi.

1. **Hendro Prasetyo, Kuntoro, Windhu Purnomo, Soenarnatalina,Merry Adriani, Bambang Wijanarko**

Analisis *Cluster* adalah sebuah proses untuk mengelompokkan satu set objek berdasarkan data yang telah similar certain karakteristik. *K-Means* adalah metode analisis cluster yang dimulai dengan menentukan jumlah cluster yang diinginkan. Bootstrap adalah teknik sampling dengan penggantian dari sampel asli.

Bootstrap digunakan untuk memperkirakan parameter berdasarkan data minimal menggunakan komputer. Metode ini berguna untuk memaksimalkan relatif beda dan variasi dalam cluster. Malnutrisi merupakan masalah utama di Indonesia dan masih menjadi perhatian pada anak balita. Bayi dengan gizi buruk akan memiliki tingkat kematian yang lebih tinggi. Tujuan dari penelitian ini wasto menilai akurasi *K-Means* dan metode *Bootstrap K-Means* untuk pengelompokan status gizi anak undersfive yang crosstabulated dengan status gizi anak berdasarkan WHO-2005 di Puskesmas Ajung, Jember. Variabel dalam penelitian ini adalah status gizi berdasarkan kriteria WHO 2005 tolok ukur sebagai standar, prosentase dan berat. Ini adalah penelitian non-reaktif, menggunakan data sekunder di Ajung Puskesmas, tanpa interaksi langsung dengan subjek. Penelitian ini menyimpulkan bahwa total tingkat akurasi (TAR) dan *Total Error Rate* (TER) untuk menentukan status gizi metode *K-Means* adalah TAR=0,9 dan , TER=0,1 ; Bootstrap KMeans metode (B=25) TAR=0,925 dan TER=0,075 ; *Bootsstrap K-Means* metode (B=50) TAR=0,9417 , TER=0,0583, dan *Bootstrap K-Means* (B=75) TAR=0,9583 dan TER=0,0417 setelah crosstabulated

1. **Joanna Ardhyanti Mita Nugraha, Yupie Kusumawati**

Persediaan obat merupakan salah satu faktor penting untuk kelangsungan proses pada Puskesmas. Belum adanya metode baku yang diterapkan merupakan masalah yang sering terjadi. Karena pada umumnya persediaan obat pada Puskesmas hanya berdasarkan pada pengalaman sebelumnya. Yaitu hanya dengan memeriksa persediaan obat yang hampir habis, kemudian membelinya.

Hal ini kurang maksimal jika suatu waktu membutuhkan obat dalam jumlah yang besar dan ternyata stok habis. Penerapan data mining dapat membantu untuk menganalisa pemakaian obat, untuk mendapatkan informasi yang dapat digunakan untuk perencanaan dan pengendalian persediaan obat. Metode clustering merupakan salah satu fungsional dari data mining yang merupakan pengelompokkan item data ke dalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial.

Algoritma K-Means merupakan algoritma yang paling popular dan banyak digunakan dalam penggunaan metode clustering data mining. Hasil yang diperoleh berupa informasi jenis obat dengan pemakaian cepat setiap bulannya yang diambil dari data 3 tahun terakhir, yang nantinya dapat digunakan sebagai metode pengendalian persediaan obat yang lebih efektif dan efisien pada Puskesmas Pandanaran.

1. **Taslimdan Fajrizal**

Melalui program jaminan kesehatan pemerintah berupaya terus menjamin kesehatan bagi masyarakat melalui puskesmas puskesmas atau balai pengobatan. Salah satu komponen yang sangat penting pada puskesmas maupun balai pengobatan adalah masalah ketersediaan obat. Ketersediaan obat harus dikelola secara baik untuk menjamin obat yang dibutuhkan oleh masyarakat selalu tersedia dengan jumlah yang cukup dan memadai.

Clusterisasi pada data mining dapat digunakan untuk menganalisa pemakaian obat yang terjadi selama ini pada sebuah puskesmas untuk digunakan sebagai salah satu alat bantu penunjang keputusan bagi pihak puskesmas untuk mengajukan permintaan obat pada periode yang akan datang. Hasil dari penelitian ini dapat mengelompokkan tingkat pemakaian obat pada apotik puskesmas Rumbai Bukit Pekanbaru

1. **Aline Embun Pramadhani dan Tedy Setiadi**

Pengolahan data yang disimpan tidak hanya bisa disimpan saja, tetapi bisa dijadikan suatu representasi pengetahuan dikemudian hari. Data pasien yang terkena Ispa digunakan untuk merepresentasikan pengetahuan dari gejala penyakit ispa sebelumnya, yang mana diderita oleh pasien di klinik pengobatan Dharma Husada. Dengan demikian, perlu adanya suatu klasifikasi penyakit yang paling banyak diderita pasien di klinik Dharma Husada.

Klasifikasi pada penelitian ini bertujuan untuk membentuk suatu model pohon keputusan untuk memprediksi penyakit pasien dan melihat variable yang paling mempengaruhi penyakit pasien dengan kategori ispa. Obyek penelitian ini adalah hasil data pasien, status imunisasi, jenis kelamin, usia dan kriteria gizi pada tahun 2012 yaitu sebanyak 200 pasien terkena ispa. Data yang digunakan adalah data pemeriksaan pasien oleh dokter untuk kemudian dapat dinyatakan terkena penyakit ispa.

Variable bebas atau variable *input* (predictor) pada penelitian ini adalah pemeriksaan awal pasien yaitu meliputi kategori status imuniasasi, jenis kelamin, usia dan kriteria gizi. Sedangkan kondisi pasien yang diprediksi yaitu terkena ispa atau tidak merupakan *output* (variable tidak bebas). Tools yang digunakan untuk membuat aplikasi penerapan data mining adalah dengan VB.Net dengan pengolahan database SQL server. Pengujian terhadap pembentukan pohon keputusan menggunakan uji Chi Square untuk mengetahui nilai hipotesis hitung yang dibandingkan dengan tabel chi square. Dari pohon keputusan yang terbentuk dari 200 data pasien maka dapat diketahui bahwa jenis kelamin tidak berpengaruh terhadap penyakit ispa.

1. **Randi Rian Putradan Cendra Wadisman**

Penelitian ini dilakukan untuk mempelajari data mining menggunakan algoritma *K-Means* yang diimplementasikan dalam pemilihan pelanggan potensial di MC Laundry. data dikumpulkan melalui observasi dan interview yang dilakukan kepada pemilik perusahaan yang bersangkutan. Selanjutnya data dianalisa untuk menentukan data yang dibutuhkan dalam proses penentuan pelanggan potensial. Dari eksperimen yang dilakukan dengan menggunakan algoritma *K-Means* serta menggunakan *Software Tanagra* dapat memberikan hasil yang akurat dalam mengelompokkan pelanggan potensial.

1. **Pengertian Algoritma**

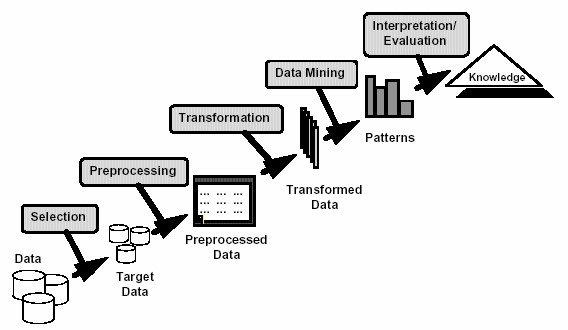
Pengertian algoritma menurut Thomas H. Cormen dalam Andi Nugroho (2011:11),”Algoritma adalah prosedur komputasi yang mengembalikan beberapa nilai atau kumpulan nilai sebagai input kemudian diproses sebagai output sehingga algoritma merupakan urutan langkah komputasi yang mengubah inout menjadi output[1].

Menurut Donal E. Knuth dalam Andi Nugroho(2011:10),”Algoritma adalah sekumpulan aturan-aturan berhingga yang memberikan sederetan operasi-operasi untuk menyelesaikan suatu jenis masalah yang khusus[2].

1. **Pengertian Data Mining**

Menurut Han dan Kamber (2011, p36), “*data mining* adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar” [3].

Lalu menurut Connolly dan Begg,(2010), “*Data mining* adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting”[4].

1. **Arsitektur Data Mining**

**Gambar 2.1 Aritektur Data Mining**

1. **Pengertian K-Means**

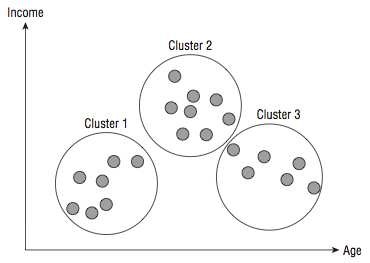
Menurut Majd Latah (2016:2), “Salah satu metode pengelompokan yang paling populer adalah k-means algoritma pengelompokan. Dalam K berarti pengamatan akan dipartisi menjadi k cluster dengan cara yang masing-masing pengamatan Milik cluster dengan mean terdekat”.

Pertama k diinisialisasi berdasarkan jumlah cluster yang diinginkan. Setiap titik data adalah ditugaskan ke pusat massa terdekat dan set poin yang ditugaskan ke centroid disebut cluster. Setiap cluster centroid adalah dimutakhirkan berdasarkan pada poin yang ditetapkan untuk cluster. Itu proses akan diulang sampai centroid tetap sama atau tidak ada titik yang mengubah cluster. Dalam algoritma ini sebagian besar Euclidean jarak digunakan untuk menemukan jarak antara titik data dan centroid. Kelemahan utama dari algoritma K-means adalah kualitas hasil pengelompokan sangat tergantung pada acak pemilihan centroid awal. Untuk menjalankan berbeda itu memberi cluster berbeda untuk input data yang sama.

Menurut Survei Chauhan A, Mishra G, dan Kumar G (2011).”Clustering digunakan untuk membuat grup (cluster) dari data sehingga dapat dengan mudah menemukan data yang diperlukan.

Clustering adalah klasifikasi benda yang mirip menjadi beberapa kelompok yang berbeda, biasanya diterapkan dalam analisis data statistik yang dapat digunakan di berbagai bidang, misalnya, pembelajaran mesin, data penambangan, pengenalan pola, analisis gambar dan bioinformatika”.

Fungsi dari *clustering* sendiri adalah untuk mencari pengelompokkan atribut ke dalam segmentasi-segmentasi berdasarkan similaritas. Berikut adalah gambaran dari clustering :



**Gambar 2.2 *Clustering***

1. **Langkah - Langkah K-Means**
   * + - 1. Pilih jumlah *cluster* k.
         2. Inisialisasi pusat *cluster k*. Secara umum pusatpusat cluster diberi nilai awal dengan angkaangka random.
         3. 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*. Untuk menghiutng jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean sesuai dengan Persamaan :

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 data ke-k

* + - * 1. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu.
        2. Periksa setiap objek pemakai pusat *cluster* yang baru. Jika pusat cluster tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah 3 sampai pusat *cluster* tidak berubah lagi.

1. **Pengertian Puskesmas**

Menurut Depkes (2011), “Pusat Kesehatan Masyarakat (Puskesmas) adalah salah satu sarana pelayanan kesehatan masyarakat yang amat penting di Indonesia. Puskesmas adalah unit pelaksana teknis dinas kabupaten/kota yang bertanggungjawab menyelenggarakan pembangunan kesehatan di suatau wilayah kerja”.

Menurut Depkes (2014), “Pusat Kesehatan Masyarakat yang selanjutnya disebut Puskesmas adalah fasilitas pelayanan kesehatan yang menyelenggarakan upaya kesehatan masyarakat dan upaya kesehatan perseorangan tingkat pertama, dengan lebih mengutamakan upaya promotif dan preventif, untuk mencapai derajat kesehatan masyarakat yang setinggi-tingginya di wilayah kerjanya.”.

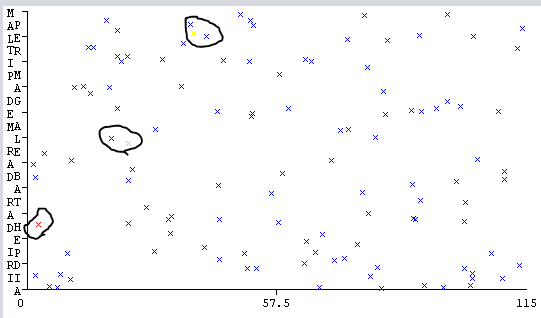
**BAB III**

**METODE PENELITIAN**

**3. Metode Penelitian yang Digunakan**

1. ***Simple Kmeans***

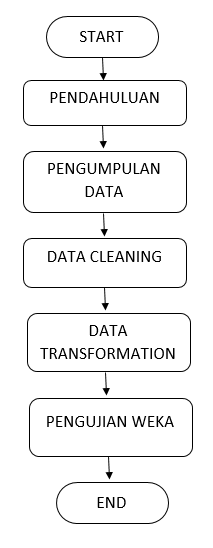
Penelitian ini menggunakan *Simple Kmeans* adalah salah satu algoritma cluster yang bertujuan untuk membagi data menjadi beberapa kelompok dan algoritma ini menerima data tanpa lebel kelas. Pada algoritma ini komputer akan mengelompokkan data tanpa mengetahui target kelasnya, berikut adalah contoh algoritma Kmeans:



Gambar 3.1 Algoritma *Kmeans*

1. **Desain Penelitian**

Metode penelitian merupakan cara ilmiah untuk mendapatkan data dengan tujuan dan tujuan dan kegunaan tertentu. Untuk mencapai tujuan tersebut diperlukan suatu metode yang relevan dengan tujuan yang ingin dicapai. Metode penelitian bertujuan untuk menguraikan seluruh kegiatan yang dilaksanakan selama kegiatan penelitian berlangsung adapun alur tahapan pada penelitian ini dapat dilihat pada gambar berikut :



Gambar 3.2 Desain Penelitian

1. Pendahuluan

Merupakan tahap awal yang dilakukan untuk mengetahui metode apa yang akan digunakan dalam menyelesaikan permasalahan yang akan diteliti dan mendapatkan referensi dasar yang kuat dalam menerapkan metode tersebut. Studi pustaka merupakan salah satu cara yang dilakukan untuk melihat dan membandingkan penelitian yang sudah ada sebelumnya.

1. Pengumpulan Data

Tahap ini merupakan langkah penelitian yang berfungsi untuk memperoleh informasi atau data yang dibutuhkan dalam penelitian, data yang dimaksudkan adalah data pasien puskesmas donomulyo, Ada dua pendekatan yang dilakukan peneliti untuk memperoleh informasi dan data yaitu :

1. Studi Pustaka

Merupakan suatu metode yang dilakukan untuk memperoleh dan mengumpulkan informasi dan data dari referensi terkait. Referensi ini dapat berupa jurnal, paper , artikel atau tulisan penelitian yang membahas kasus yang serupa dengan kasus dalam penelitia ini.

1. Wawancara

Merupakan metode dilakukan dalam bentuk bertanya dan berdiskusi dalam menyelesaikan permasalahan yang ada didalam penelitian ini dengan orang yang terkait dalam proses bisnis yang berada pada puskesmas donomulyo.

1. Pembersihan Data (*Data Cleaning*)

Pembersihan data yang dimaksud adalah menghapus data yang tidak sesuai dengan kriteria, contoh data yang tidak sesuai dengan kriteria yaitu data yang kosong. Tahap ini digunakan agar data mentah nantinya dapat diolah dalam data mining dengan algoritma clustering.

Tahap pembersihan data disini adalah membuang data yang kosong dan memastikn bahwa data tersebut relevan serta terkait satu sama lain, berikut adalah contoh data sebelum dan sesudah masuk tahapan data *Cleaning*:

Gambar 3.3 Data Sebelum Dilakukan *Data Cleaning*

Terlihat diatas ada data yang tidak lengkap sebanyak 2 data dan data ini berada pada field alamat nantinya akan dihapus guna membuat data menjadi data yang relevan, berikut contoh data setelah dilakukan *data cleaning*:



Gambar 3.4 Data Setelah Dilakukan *Data Cleaning*

1. Mengubah Data Mentah Menjadi Dataset (*Data Transformasi*).

Data mentah yang telah melalui tahap cleaning atau dibersihkan akan dijadikan sebagai dataset sebelum masuk pada tahap selanjutnya data mentah harus diubah kedalam bentuk yang sesuai dengan karakteristik data mining mulai dari mengubah format menjadi .csv.

Sebelum dilakukan *transformasi* data adapun atribut yang dimiliki adalah no, nama, usia, diagnosa, jenis kelamin dan alamat yang dapat dilihat pada gambar dibawah :



Gambar 3.5 Data Sebelum dilakukan *Transformasi* data

Setelah dilakukan *transformasi* data atribut yang dimiliki adalah no, usia, jenis kelamin dan alamat, transformasi data pada jenis kelamin yaitu jika laki – laki maka di wakili dengan angka 1 sementara perempuan diwakili angka 1. Dapat dilihat pada gambar dibawah :



Gambar 3.5 Data Sesudah dilakukan *Transformasi* data

1. Pengujuan data kedalam aplikasi WEKA

WEKA atau kepanjangan dari *Waikato Environment for Knowledge Analysis* merupakan suatu aplikasi yang digunakan untuk menganalisis data mining. Didalam aplikasi WEKA terdapat banyak metode yang dapat digunakan termasuk Kmeans, dalam aplikasi ini dataset yang telah masuk tahap – tahap diatas dimasukan ke aplikasi WEKA guna mengetahui pola dari suatu data dan dapat melihat data tersebut termasuk didalam kelompok yang mana.

Aplikasi ini mampu mengelompokkan data dengan cara mengvisualisasikan data data yang berkelompok dapat juga mengvisualisasikan menggunakan grafik.

1. **Analisis Kebutuhan**
2. Analisis Kebutuhan Perangkat Keras (Hardware)

Adapun kebutuhan perangkat keras yang dibutuhkan dalam penelitian ini berupa Personal Laptop dengan spesifikasi berikut :

1. RAM 2 GB
2. Hardisk 500GB
3. Processor Intel Core i3
4. Analisis Kebutuhan Perangkat Lunak (Software

Adapun perangkat lunak yang dibutuhkan adalaha :

1. Sistem Operasi Windows 10 64-bit
2. Microsoft Office Excel 2013
3. Software WEKA (Waikato Enviroment for Knowledge Analysis).

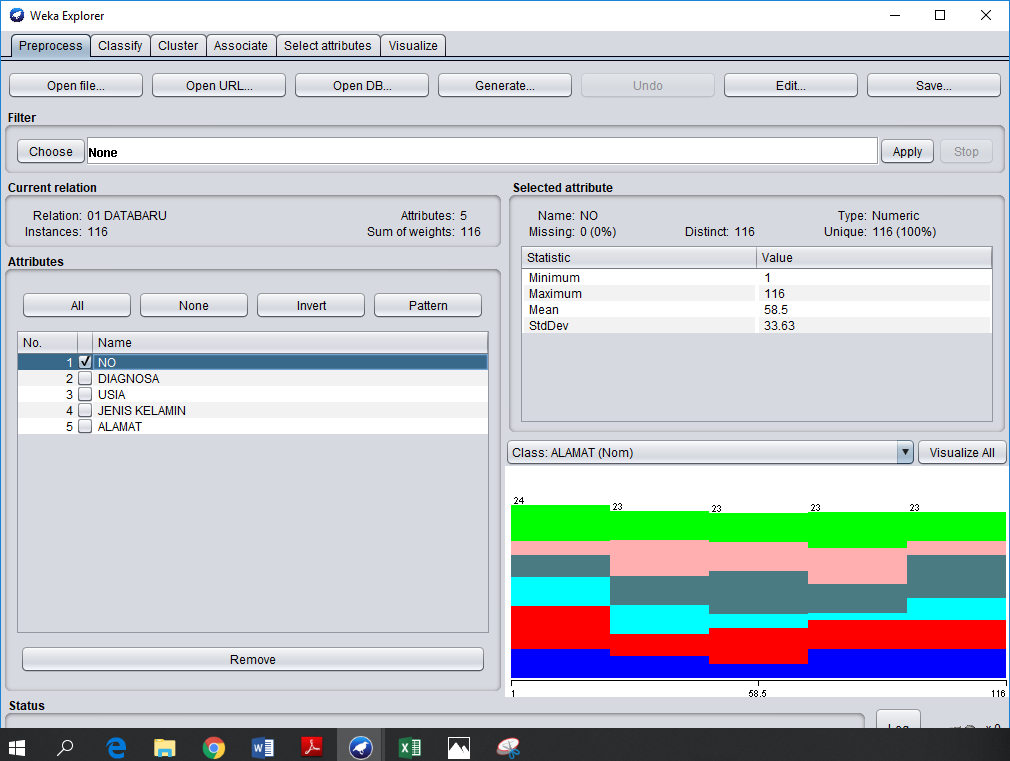
**BAB IV**

**PEMBAHASAN**

1. **Pembahasan**

Data yang telah dibersihkan dan ditransformasi akan masuk pada pengujian menggunakan aplikasi WEKA, adapun hasil nya adalah sebagai berikut :

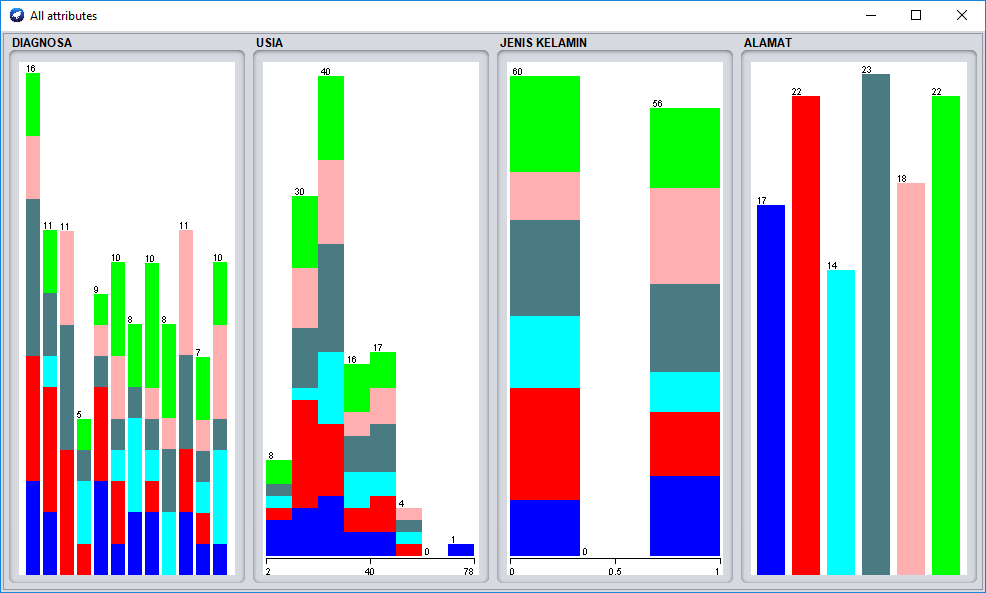
1. Memasukan data ke WEKA

Pastikan data yang akan dimasukan dalam format csv dan tiap data dipisahkan dengan koma (,) serta hapus field yang tidak dibutuhkan dalam pengujian ini yaitu field no.

Gambar 4.1 Pengujian dataset kedalam Aplikasi WEKA

1. Tampilan grafik

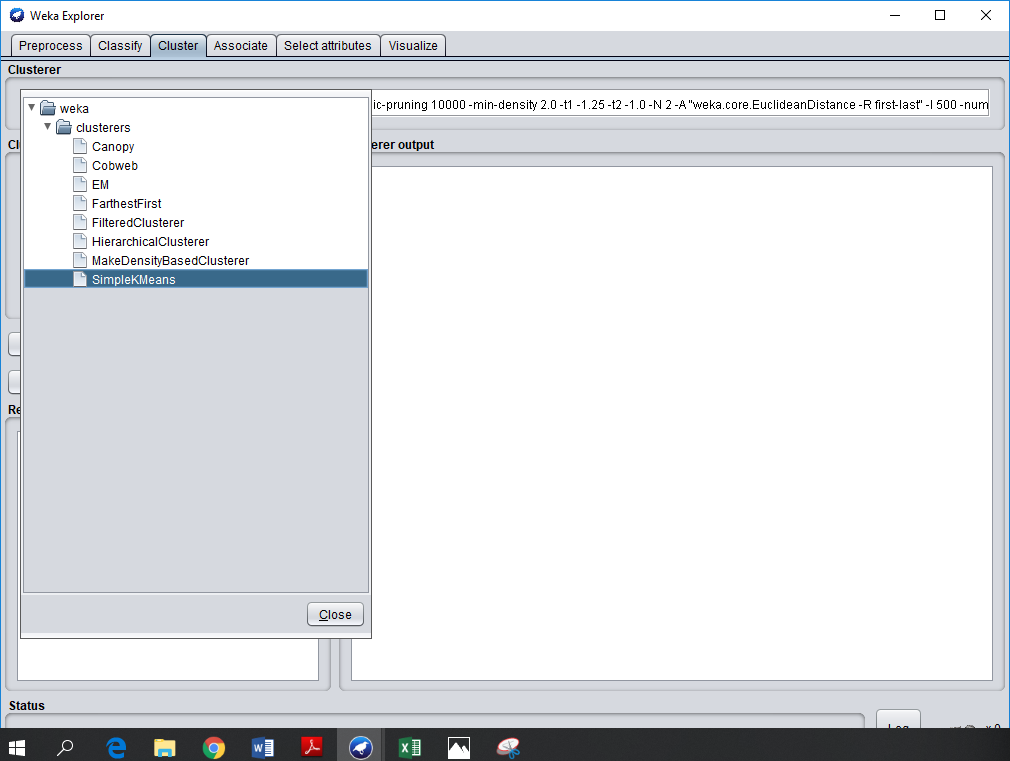
Grafik dibawah menggambarkan kriteria dari data pasien berupa diagnosa, usia, jenis kelamin dan alamat. Seluruh data divisualisasikan dengan membedakan berdasarkan warna.



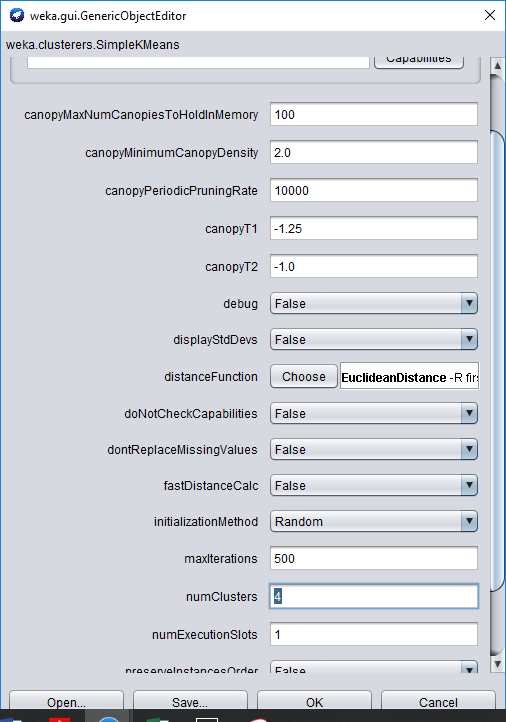
Gambar 4.2 Grafik Dataset

1. Proses Clustering

Pada proses ini data yang telah kita masukan akan siap diuji kedalam aplikasi, sebelum melakukan tahap tersebut langkah yang harus kita lakukan adalah memilih algoritma serta jumlah cluster yang diinginkan.

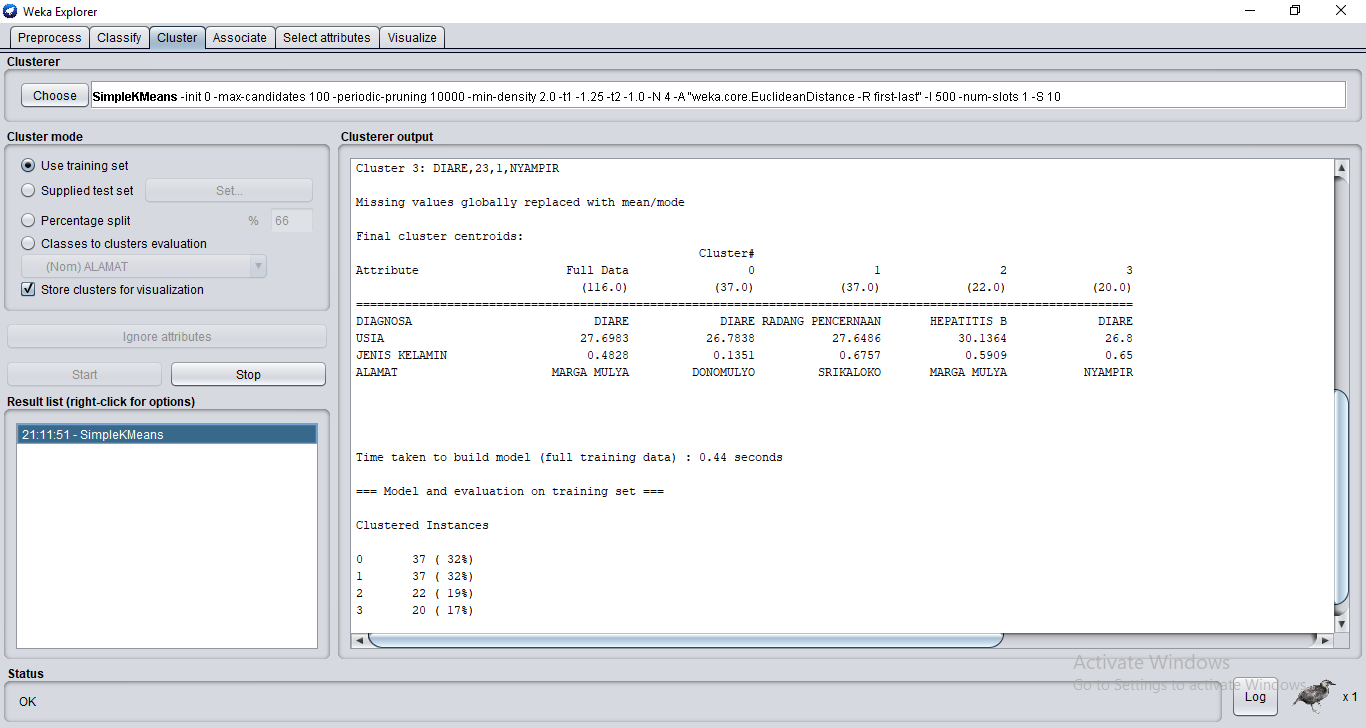
dalam pengujian ini peneliti menggunakan cluster sebanyak 4 , berikut dapat dilihat pada gambar dibawah :

Gambar 4.3 Memilih Algoritma Pada WEKA

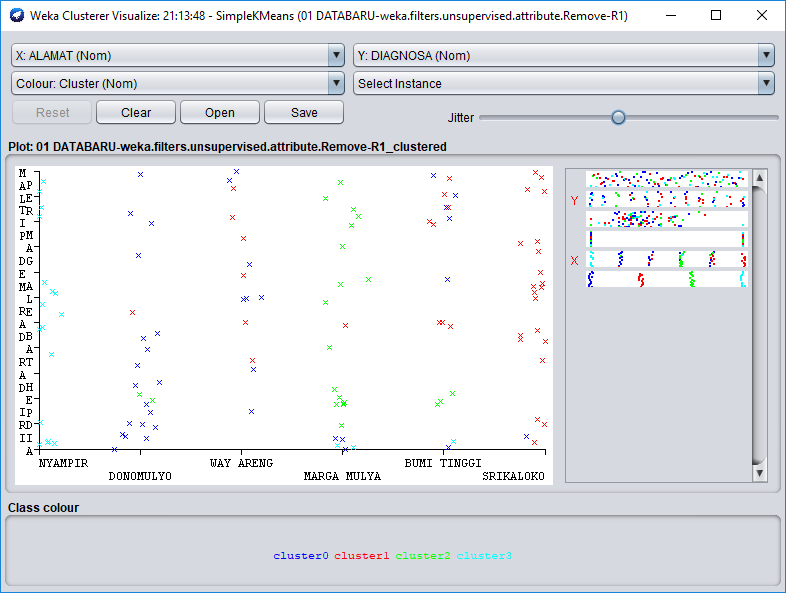
  
Gambar 4.4 Memasukan Jumlah *Clusters*

1. Hasil

Berikut adalah hasil dari clustering yang dihasilkan dari proses pengujian dengan WEKA dengan 4 jumlah cluster dengan hasil sebagai berikut :



Gambar 4.5 Hasil Kmeans

Pada gambar diatas dapat disimpulkan bahwa hasil pada cluster 0 sebanyak 37 data dengan persentase 32%, cluster 1 sebanyak 37 data dengan persentase 32%, cluster 2 sebanyak 22 data dengan persentase 19% sementara cluster 3 sebanyak 20 data dengan persentase 17%. Berikut merupakan visualizer dari clustering dengan menggunakan sumbu X Alamat dan sumbu Y Diagnosa:

Gambar 4.6 Visualizer hasil Kmeans