

## ANALISIS DAN PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA RUMAH SAKIT UMUM DAERAH PRABUMULIH

Trisca Anggraini Putri<sup>1</sup>, Nurul Huda<sup>2</sup>

Fakultas Teknik Ilmu Komputer, Universitas Bina Darma  
Email: trisca08@gmail.ac.id<sup>1</sup>, nurul\_huda@binadarma.ac.id<sup>2</sup>

### ABSTRAK

Penyakit jantung merupakan penyakit kondisi ketika jantung mengalami gangguan yang dapat menyerang secara tiba-tiba dan menyebabkan kematian tanpa memandang usia dan jenis kelamin. Tingginya jumlah kematian akibat penyakit jantung disebabkan masih kurangnya kesadaran masyarakat untuk melakukan pengecekan kesehatan jantung secara berkala dan buruknya gaya hidup masyarakat. Pada penyakit jantung banyak mengalami gangguan, bentuk gangguan itu sendiri bisa bermacam-macam seperti pembuluh darah jantung, irama jantung, katup jantung, atau gangguan akibat bawaan lahir. Oleh karena itu dengan perkembangannya zaman dan peradaban manusia maka teknologi dan ilmu kesehatan dapat dikolaborasikan menjadi sesuatu yang dapat bermanfaat bagi manusia dan sekitarnya. Dengan melakukan analisis dan prediksi penyakit jantung. Pada penelitian ini berfokus pada dataset diagnosa medis pasien menggunakan algoritma *k-means*.

**Kata kunci:** Data Mining, Analisis, Prediksi, Algoritma K-means, Clustering

### ABSTRACT

*Heart disease is a condition when the heart experiences a disorder that can strike suddenly and cause death regardless of age and gender. The high number of deaths due to heart disease is due to the lack of public awareness to check heart health regularly and the poor lifestyle of the community. In heart disease there are many disorders, the form of the disorder itself can vary, such as heart blood vessels, heart rhythm, heart valves, or congenital disorders. Therefore, with the development of the era and human civilization, technology and health sciences can be collaborated into something that can benefit humans and their surroundings. By analyzing and predicting heart disease. This study focuses on a dataset of patient medical diagnoses using the *k-means* algorithm.*

**Keywords:** *Data Mining, Analysis, Prediction, K-means Algorithm, Clustering*

## 1. PENDAHULUAN

Pada tahun 1947 berdirilah Balai Pengobatan yang merupakan cikal bakal Rumah Sakit Umum Daerah (RSUD) Kota Prabumulih, lokasinya adalah bangunan eks kantor Marga Kapak tengah (Lokasi Lapangan Tenis Dusun Prabumulih sekarang ini). Pada tahun 1955 Balai Pengobatan tersebut dikembangkan menjadi Rumah Sakit Umum Daerah Kota Prabumulih yang berlokasi dari tahun 1955 sampai dengan akhir tahun 2008 di Jl. AK. Gani No. 41 Karang Raja III Kelurahan Tugu Kecil Prabumulih Timur. Dengan berubah status Kota Administratif Prabumulih menjadi Kota Prabumulih sebagai hasil pemekaran dari Kabupaten Muara Enim sesuai dengan Undang-undang Nomor 6 tahun 2001, maka dengan sendirinya RSUD Kota Prabumulih diserahkan ke Pemerintah Kota Prabumulih.

Pada tahun 2004 sudah mulai dibangun RSUD Kota Prabumulih di kawasan Jl. Lingkar Kelurahan Gunung Ibul Prabumulih Timur 31111 dengan luas 5 hektar. Pada tanggal 3 Desember 2008 RSUD Kota Prabumulih telah melaksanakan perpindahan pelayanan ke lokasi tersebut, dan mulai dapat memberikan Pelayanan Kesehatan kepada masyarakat Kota Prabumulih pada tanggal 6 Desember 2008 sampai dengan sekarang [5].

Penyakit jantung merupakan sebuah kondisi ketika jantung mengalami gangguan. Bentuk gangguan itu sendiri bisa bermacam-macam. Ada gangguan pada pembuluh darah jantung, irama jantung, katup jantung atau gangguan akibat bawaan lahir. Fungsi utama jantung adalah mengalirkan darah kaya oksigen ke seluruh bagian tubuh. Setelah seluruh organ tubuh menggunakan oksigen dalam darah, darah yang miskin oksigen tersebut kembali ke jantung (atrium kanan), untuk diteruskan ke ventrikel kanan melalui katup trikuspid. Sesudah darah memenuhi ventrikel kanan, katup trikuspid akan menutup guna mencegah darah kembali ke atrium kanan.

Penyakit jantung sampai saat ini merupakan salah satu penyakit yang memerlukan perhatian khusus, dimana menurut hasil riset kesehatan dasar (Riskesdas), angka penyakit jantung dan pembuluh darah semakin meningkat dari tahun ke tahun. Penyakit jantung merupakan salah satu masalah penyakit utama dan penyebab nomor satu kematian didunia. Ada beberapa penyakit jantung seperti gejala jantung, gagal jantung, jantung rematik, jantung iskemik dll. Penyebab penyakit jantung sangat bervariasi, mulai dari masalah pada pembuluh darah jantung, irama jantung, hingga bawaan lahir dan penyebab penyakit jantung berdasarkan jenisnya. Dan kurangi aktivitas, stress yang tidak ditangani dengan baik, serta kondisi medis seperti diabetes atau obesitas, dan juga dapat meningkatkan resiko penyakit jantung.

Gejala penyakit jantung sangat beragam, tergantung kepada jenis kondisi yang dialami. Sejumlah gejala yang dapat muncul adalah nyeri dada terasa seperti tertindih. Nyeri dileher, rahang, tenggorokan, punggung, dan lengan. Jantung berdebar atau detak jantung malah melambat dan sesak nafas.

Dari sumber yang saya dapatkan dari RSUD Prabumulih dalam 3 tahun terakhir pasien yang menderita penyakit jantung berjumlah 304 pasien, diantaranya laki-laki berjumlah 151 orang dan perempuan berjumlah 153 orang dan data tersebut didapatkan dari hasil penelitian yang dilakukan di RSUD Prabumulih data tersebut yang diolah dan dilaporkan kepada pihak eksekutif untuk mengetahui perkembangan rumah sakit pada waktu tertentu.

Dari permasalahan di atas untuk mengatasi banyaknya penyakit jantung yang ditimbulkan, maka dilakukannya analisis dan prediksi menggunakan Algoritma *K-Means Clustering*. Yang mana nantinya akan menentukan jenis penyakit jantung yang sering terjadi atau sering di alami, hasil dari prediksi tersebut akan membandingkan jenis penyakit jantung yang paling banyak di alami, hasil tersebut nantinya bisa dilihat dari tabel atau grafik prediksi yang sudah di analisis. Berdasarkan latar belakang di atas yang telah di uraikan maka penulis mengangkat judul “Analisis Dan Prediksi Penyakit Jantung Menggunakan Algoritma *K-Means Clustering* Pada Rumah Sakit Umum Daerah Prabumulih”.

## 2. METODOLOGI PENELITIAN

### 2.1 Penyakit Jantung

Jantung merupakan organ manusia yang berperan dalam sistem peredaran darah. Penyakit jantung adalah sebuah kondisi dimana jantung tidak dapat melaksanakan tugasnya dengan baik. Data WHO menyatakan bahwa sebanyak 7,3 juta penduduk dunia meninggal dikarenakan penyakit jantung. Meskipun penyakit jantung merupakan penyakit yang tidak menular, penyakit ini merupakan jenis penyakit yang mematikan nomor satu di dunia.

Penyakit jantung disebut juga dengan penyakit jantung koroner, penyakit ini terjadi bila darah ke otot jantung terhenti/tersumbat, sehingga mengakibatkan kerusakan berat pada jantung. Penyebab utama penyakit jantung adalah penggunaan tembakau, fisik tidak aktif, diet yang tidak sehat dan penggunaan alkohol, resiko penyakit jantung bertambah dengan meningkatkan usia, tekanan darah tinggi, mempunyai kolesterol tinggi, dan kelebihan berat badan [3].

## 2.2. Data Mining

Dalam jurnal (Ridwan and Suyono 2013) Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [8][9]. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki [10][11]. Data mining, sering juga disebut sebagai *Knowledge Discovery in Database (KDD)* [13]. KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [6][12].

### A. Metode Pelatihan

Metode Pelatihan Secara garis besar metode pelatihan yang digunakan dalam teknik teknik data mining dibedakan ke dalam dua pendekatan, yaitu :

#### a. *Unsupervised learning*

Metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*). Guru di sini adalah label dari data [14].

#### b. *Supervised learning*

Yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai *output* atau label selama proses *training* [15].

### B. Pengelompokan Data Mining

Ada beberapa teknik yang dimiliki data mining berdasarkan tugas yang bisa dilakukan, yaitu [16] :

#### a. Deskripsi

Para peneliti biasanya mencoba menemukan cara untuk mendeskripsikan pola dan trend yang tersembunyi dalam data.

#### b. Estimasi

Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numerik dari pada kategori.

#### c. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi di masa depan).

#### d. Klasifikasi

Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

#### e. Clustering

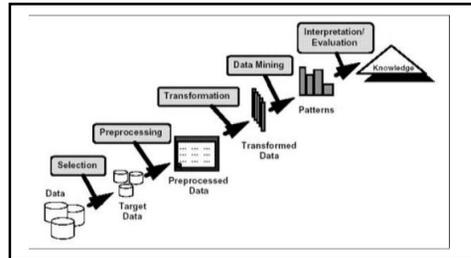
Clustering lebih ke arah pengelompokan record, pengamatan, atau kasus dalam kelas yang memiliki kemiripan. Asosiasi Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

#### f. Asosiasi

Asosiasi mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

### C. Tahap – Tahap Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada Gambar 1. Tahap-tahap tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain, dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. [4] [17].



**Gambar 1.** Tahap – Tahap Data Mining

Tahap-tahap data mining adalah sebagai berikut :

1. *Data selection*  
Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discovery in Database (KDD)* dimulai.
2. *Data hasil*  
seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas terpisah dari basis data operasional.
3. *Pre-processing / Cleaning*  
Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus *Knowledge Discovery in Database (KDD)*. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Juga dilakukan proses enrichment, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlakukan untuk *Knowledge Discovery in Database (KDD)*, seperti data atau informasi eksternal lainnya yang diperlukan.
4. *Transformation*  
*Coding* adalah proses transformasi pada data yang dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam *Knowledge Discovery in Database (KDD)* merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
5. *Data Mining*  
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* secara keseluruhan.
6. *Interpretation / Evaluation*  
Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses *Knowledge Discovery in Database (KDD)* yang disebut interpretasi. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### 2.3. Clustering

Clustering menurut Darmi and Setiawan [1] Clustering atau klasifikasi adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa group berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. Cluster adalah sekelompok atau sekumpulan objek-objek data yang similar satu sama lain dalam cluster yang sama dan dissimilar. terhadap objek-objek yang berbeda cluster .

Objek akan dikelompokkan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan yang lainnya. Objek-objek dikelompokkan berdasarkan prinsip memaksimalkan kesamaan objek pada cluster yang sama dan memaksimalkan ketidaksamaan pada cluster yang berbeda. Kesamaan objek biasanya diperoleh dari nilai-nilai atribut yang menjelaskan objek data, sehingga objek-objek data biasanya dipresentasikan sebagai sebuah titik dalam ruang multidimensi.

Dengan menggunakan clustering ini, kita dapat mengkategorikan daerah yang padat, menemukan pola-pola distribusi secara keseluruhan, dan menemukan keterkaitan yang menarik antara atribut data. Dalam data mining, usaha difokuskan pada metode-metode penemuan untuk cluster pada basis data berukuran besar secara efektif dan efisien . Beberapa kebutuhan clustering dalam data mining meliputi skalabilitas, kemampuan untuk menangani tipe atribut yang berbeda mampu menangani dimensionalitas yang tinggi, menangani data yang mempunyai noise, dan dapat diterjemahkan dengan mudah.

Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan objektif function yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi dalam suatu cluster. Dan meminimalisasikan variasi antar *cluster*. Secara garis besar, terdapat beberapa metode klasifikasi data. Pemilihan metode clustering tergantung pada tipe data dan tujuan clustering itu sendiri.

### 2.4. Algoritma *K-Means Clustering*

*K-Means* merupakan salah satu metode data clustering non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbedadikelompokkan ke dalam kelompok yang lainnya.

*K-Means* adalah metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numeric. Algoritma *K-Means* termasuk partitioning clustering yang memisahkan data ke *k* daerah bagian yang terpisah . Algoritma *K-Means* sangat terkenal karena kemudahannya dan kemampuannya untuk mengcluster data yang besar dan data outlier dengan sangat cepat. Dalam algoritma *K-Means*, setiap data harus termasuk ke cluster tertentu dan bisa dimungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah kecluster lainnya.

Pada dasarnya penggunaan algoritma dalam melakukan proses clustering tergantung dari data yang ada dan konklusi yang ingin dicapai. Untuk itu digunakan algoritma *K-Means* yang didalamnya membuat aturan sebagai berikut:

1. Jumlah *Cluster* perlu diinputkan
2. Hanya memiliki atribut bertipe numeric.

Algoritma *K-Means* merupakan metode *nonhierarki* yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat *cluster* awal . Pada tahap ini pusat cluster dipilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat cluster yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap *cluster*. Posisi pusat *cluster* akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap pusat dan terakhir akan terbentuk posisi pusat cluster yang baru. Algoritma *K-Means cluster* pada dasarnya melakukan dua proses, yakni proses pendeteksian lokasi pusat tiap *cluster* dan proses pencarian anggota dari tiap-tiap cluster. Cara kerja algoritma *K-Means*:

1. Masukkan data yang akan dikluster.
2. Tentukan jumlah kluster.

3. Ambil sebarang data sebanyak jumlah kluster secara acak sebagai pusat kluster (sentroid).
4. Hitung jarak antara data dengan pusat kluster, dengan menggunakan persamaan.

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

Dimana :

$D(i, j)$  = jarak data ke  $i$  ke pusat kluster  $j$

$X_{ki}$  = data ke  $i$  pada atribut ke  $k$

$X_{kj}$  = titik pusat ke  $j$  pada atribut ke  $k$

5. Hitung kembali pusat kluster dengan keanggotaan kluster yang baru
6. Jika pusat kluster tidak berubah maka proses kluster telah selesai, jika belum maka ulangi langkah ke (4) sampai pusat kluster tidak berubah lagi.

Menurut Irfiani and Rani [2] metode *K-Means* memiliki karakteristik sebagai berikut:

1. *K-Means* merupakan metode pengelompokan yang sederhana dan dapat digunakan dengan mudah.
2. Pada jenis data set tertentu, *K-Means* tidak dapat melakukan segmentasi data dengan baik di mana hasil segmentasi tidak dapat menentukan pola kelompok yang mewakili karakteristik bentuk alami data.
3. *K-Means* bias menalami masalah ketika mengelompokkan data yang mengandung outlier.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Data Penyakit jantung

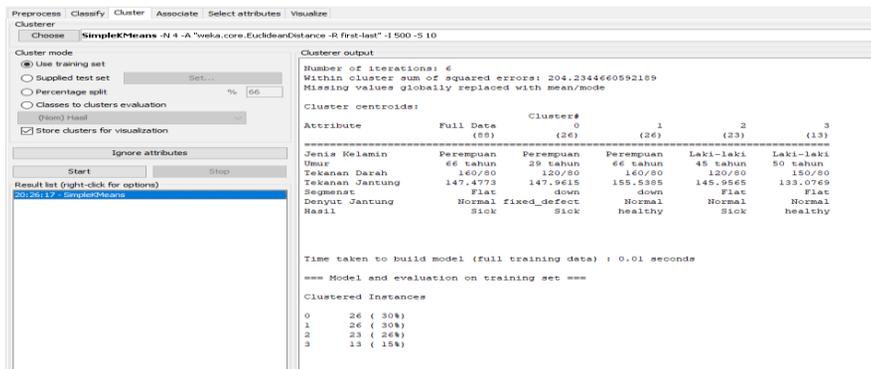
Dalam kasus ini, penulis menggunakan sample data dari penyakit jantung periode 2016 – 2018 dan data tersebut berjumlah 304 pasien. Data yang didapat adalah sebagai berikut :

**Tabel 1. Rekapitulasi Penyakit Jantung pada Rumah Sakit Umum Daerah Prabumulih Tahun 2016 – 2018**

Umur	Jenis Kelamin	Tekanan Darah	Tekanan Jantung	Segmenst Tekanan Jantung	Denyut Jantung	Hasil
6 tahun	Laki – laki	120 / 70	108	Down	Normal	Sick
27 tahun	Perempuan	100 / 90	129	Down	Normal	Sick
29 tahun	Laki – laki	120 / 90	150	Flat	Fixed_defect	Sick
25 tahun	Laki – laki	130 / 80	145	Flat	Fixed_defect	Healthy
45 tahun	Laki – laki	140 / 80	172	Flat	Normal	Healthy
31 hari	Perempuan	110 / 70	150	Down	Normal	Sick
20 tahun	Perempuan	160 / 80	160	Down	Normal	Sick
26 tahun	Laki – laki	120 / 90	129	Flat	Fixed_defect	Healthy
40 tahun	Perempuan	100 / 90	187	Down	Fixed_defect	Healthy
29 tahun	Perempuan	130 / 80	147	Down	Normal	Healthy
60 tahun	Perempuan	120 / 70	129	Flat	Fixed_defect	Sick
65 tahun	Laki – laki	140 / 80	147	Flat	Normal	Sick
66 tahun	Laki – laki	160 / 80	147	Flat	Normal	Sick
59 tahun	Perempuan	100/ 90	108	Down	Normal	Sick
13 tahun	Laki – laki	110 / 70	160	Down	Fixed_defect	Sick

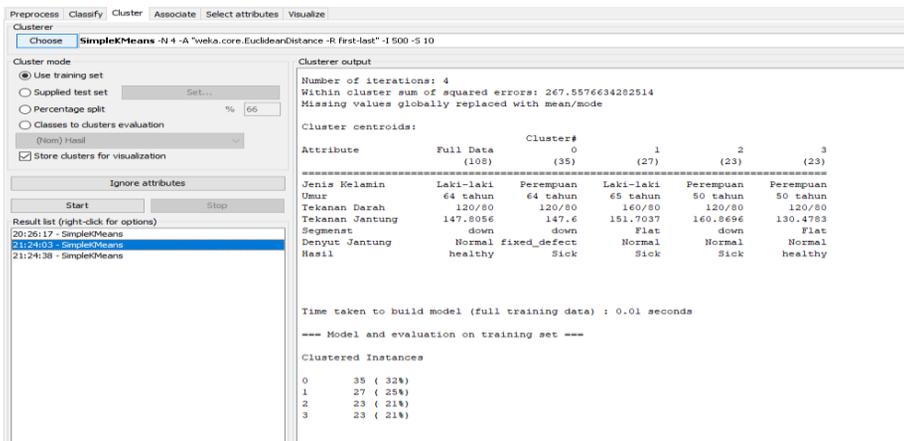
### 3.2 Hasil Analisis dan Prediksi Menggunakan Weka

Dalam jurnal Saputra, Adji, and Permanasari [7] WEKA adalah sebuah aplikasi yang dipakai untuk pembelajaran mesin (*machine learning*) yang dimaksudkan untuk membantu dalam penerapan teknik pembelajaran mesin keberbagai masalah yang ada dalam dunia nyata, khususnya dalam penelitian ini berhubungan dengan politik dan manajemen reputasi. Pembelajaran yang ada dalam WEKA meliputi visualisasi hasil, manipulasi data, hubungan antar basis data, *cross-validation*, dan perbandingan set aturan untuk melengkapi penelitian yang berhubungan dengan pembelajaran mesin dasar. Selain itu, WEKA menyediakan lingkungan untuk keperluan klasifikasi, *regresi*, *clustering*, dan fitur seleksi. Dengan WEKA, dapat diketahui akurasi dari mulai preProcessing hingga hasil akhir dengan metode tertentu dengan cepat dibandingkan perhitungan manual.



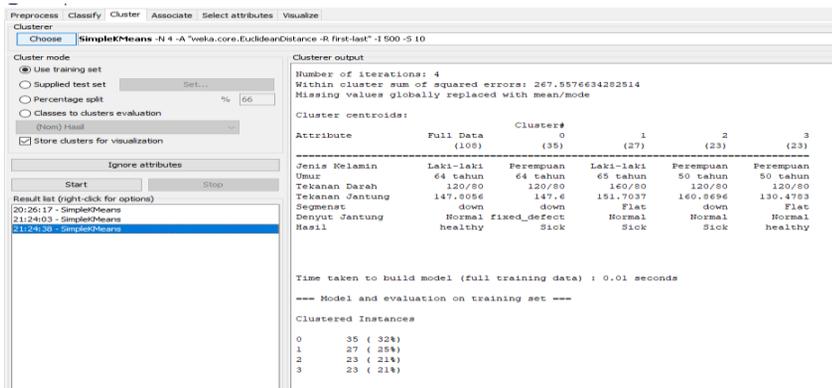
Gambar 2. Hasil prediksi pada tahun 2016

Pada tahap awal algoritma akan membentuk centroid terlebih dahulu. Centroid merupakan nilai pusat atau nilai acuan bagi data lain. Data yang di proses berjumlah 7 atribut yaitu jenis kelamin, umur, tekanan darah, tekanan jantung, segment, denyut jantung dan hasil. Selanjutnya di jelaskan bahwa aplikasi weka otomatis memilih secara acak random. Nilai pusat atau nilai acuan untuk mengelompokan data pasien peyakit jantung ke dalam cluster yang telah di tentukan agara dapat dilihat pada pola datanya. Sehingga struktur centroid akan bernilai 4. Pada kode program di atas bisa dilihat bahwa pada cluster 0 mengambil pola atribut jenis kelamin, atribut,perempuan sick dengan jumlah 88 pasien dan atribut angka sick sebanyak 26.



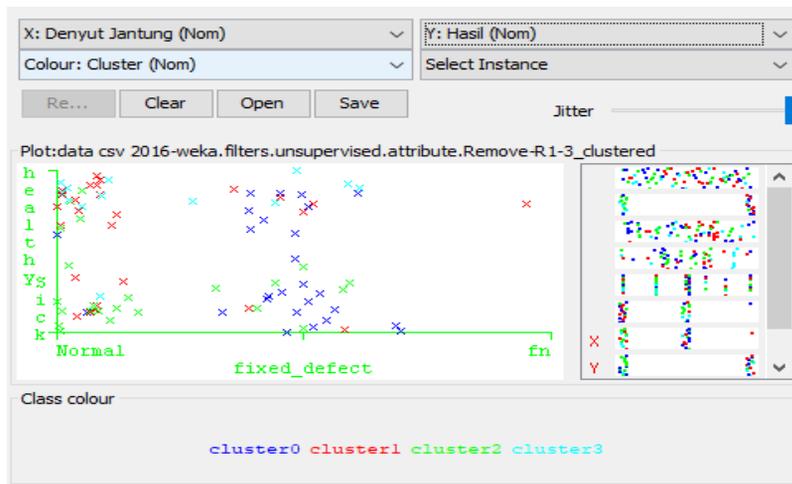
Gambar 3. Hasil prediksi pada tahun 2017

Sama seperti penjelasan di kode program 2016 pada kode program di atas hasil proses output bisa dilakukan pada aplikasi weka akan membentuk centroid yang baru dan jelas berbeda dari sebelumnya yaitu dari tahap inialisasi. Setelah di lakukan kode program *output*, selanjutnya hasil dari *cluster* data set penyakit jantung pada tahun 2017 sebanyak 35 pasien yang terkena penyakit jantung.



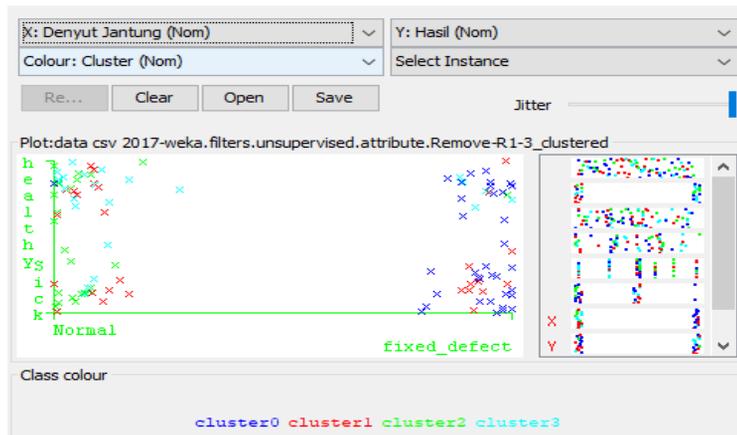
Gambar 4. Hasil prediksi pada tahun 2018

Pada kode program di atas bisa dilihat bahwa pada hasil proses *output* bisa dilihat pada aplikasi weka akan membentuk *centroid* yang baru dan jelas berbeda dari sebelumnya yaitu dari tahap inialisasi. Setelah di lakukan kode program *output*, selanjutnya hasil dari *cluster* data set penyakit jantung pada tahun 2018 menggunakan metode k-means sebanyak 35 pasien yang terkena penyakit jantung.



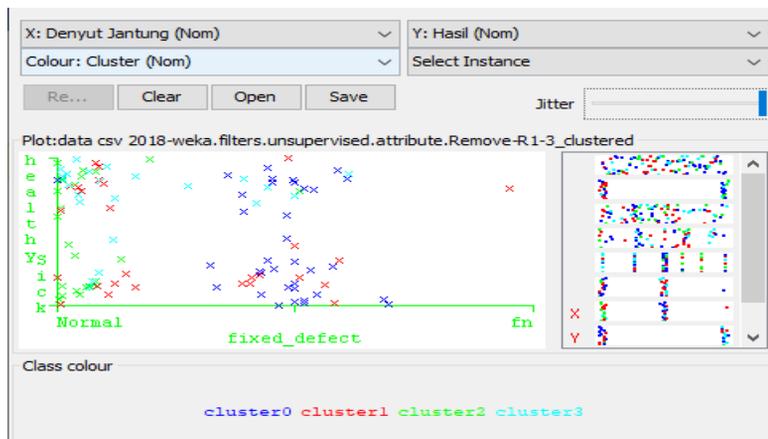
Gambar 5. Visualize Tahun 2016

Visualize tahun 2016 pada titik biru dan merah adalah visualisasi denyut jantung dan predicted hasil pada tahun 2016 dan jumlah titik biru yang paling banyak ialah predicted hasil pada penyakit jantung.



Gambar 6. Visualize 2017

Visualize pada tahun 2017 dititik biru dan merah adalah visualisasi denyut jantung dan predicted hasil pada tahun 2017 dan jumlah titik biru yang lebih banyak ialah predicted hasil.



Gambar 7. Visualisasi 2018

Visualize pada tahun 2018 dititik biru dan merah adalah visualisasi denyut jantung dan predicted hasil pada tahun 2018 dan jumlah titik biru yang lebih banyak ialah predicted hasil.

Table. 2 Hasil Prediksi Penyakit Jantung Menggunakan Weka dari Tahun 2016-2018

Tahun	Laki-laki (sick)	Laki-laki (healty)	Perempuan (sick)	Perempuan (healty)
2016	26	26	23	13
2017	35	27	23	23
2018	35	27	23	23
<b>Total</b>	72,6 %	58 %	53,6 %	43,6 %

#### 4. KESIMPULAN

Dari penelitian yang telah dilakukan menggunakan algoritma *K-Means*. Menggunakan data pasien penyakit jantung. Untuk prediksi ketepatan penyakit jantung menggunakan weka V3.6.13. Hasil penelitian ini ialah untuk yang terkena penyakit jantung berjenis kelamin laki-laki = 25%, untuk jenis kelamin perempuan yang terkena penyakit jantung = 25,33 % , dan untuk jenis kelamin laki-laki yang sembuh dari penyakit jantung = 25,33 % , untuk jenis kelamin perempuan yang sembuh dari penyakit jantung = 26 % . Dengan demikian algoritma *K-Means* adalah algoritma yang terbaik untuk memecahkan masalah dalam prediksi penyakit jantung.

#### DAFTAR PUSTAKA

- [1] Darmi, Yulia, and Agus Setiawan. 2016. *Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk*. 12(2):10.
- [2] Irfiani, Ani, and Siti Sulistia Rani. 2018. *Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita*. Jurnal Sistem dan Teknologi Informasi (JUSTIN) 6(4):161.
- [3] Lestari, Mei. 2014. *Penerapan algoritma Klasifikasi Nearest Neighbor (K-Nn) Untuk Mendeteksi Penyakit Jantung*. 6.
- [4] Mardi, Yuli. n.d. *Data Mining : Klasifikasi Menggunakan Algoritma C4.5*. 7.
- [5] Mindari, Monika, and Muhammad Nasir. n.d. *Penerapan Document Oriented Database (Nosql) Pada Sistem Informasi Rumah Sakit Umum Daerah (RSUD) Prabumulih*. 8.
- [6] Ridwan, Mujib, Hadi Suyono, & M. Sarosa. *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. Jurnal EECCIS Vol.7. No.1 tahun 2013. (Diakses <https://jurnaleeccis.ub.ac.id/index.php/eccis/article/view/204>).
- [7] Saputra, Nurirwan, Teguh Bharata Adji, and Adhistya Erna Permanasari. 2015. *Analisis sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi dan Stemming Menggunakan Metode Naive Bayes dan Svm*. 5:12.
- [8] Wanto, A., Siregar, M.N.H., Windarto, A.P., Hartama, D., Ginantra, N.L.W.S.R., Napitupulu, D., Negara, E.S., Lubis, M.R., Dewi, S.V. and Prianto, C., 2020. *Data Mining: Algoritma dan Implementasi*. Yayasan Kita Menulis
- [9] Ginantra, N.L.W.S.R., Arifah, F.N., Wijaya, A.H., Septarini, R.S., Ahmad, N., Ardiana, D.P.Y., Effendy, F., Iskandar, A., Hazriani, H., Sari, I.Y. and Gustiana, Z., 2021. *Data Mining dan Penerapan Algoritma*. Yayasan Kita Menulis
- [10] Andryani, R., Negara, E.S. and Triadi, D., 2019. *Social Media Analytics: Data Utilization of Social Media for Research*. Journal of Information Systems and Informatics, 1(2), pp.193-205.
- [11] Kaunang, F.J., Karim, A., Simarmata, J., Iskandar, A., Ardiana, D.P.Y., Septarini, R.S., Negara, E.S., Hazriani, H. and Widayastuti, R.D., 2021. *Konsep Teknologi Informasi*. Yayasan Kita Menulis.
- [12] Nurhachita, N. and Negara, E.S., 2020. *A Comparison Between Naïve Bayes and The K-Means Clustering Algorithm for The Application of Data Mining on The Admission of New Students*. Jurnal Intelektualita: Keislaman, Sosial dan Sains, 9(1), pp.51-62.

- [13] Nurhachita, N. and Negara, E.S., 2021. A comparison between deep learning, naïve bayes and random forest for the application of data mining on the admission of new students. *IAES International Journal of Artificial Intelligence*, 10(2), p.324.
- [14] Syaputra, R., Syamsuar, D. and Negara, E.S., 2021, February. Multiple Smile Detection Using Histogram of Oriented Gradient and Support Vector Machine Methods. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1071, No. 1, p. 012027). IOP Publishing.
- [15] Amanda, R. and Negara, E.S., 2020. Analysis and Implementation Machine Learning for YouTube Data Classification by Comparing the Performance of Classification Algorithms. *Jurnal Online Informatika*, 5(1), pp.61-72.
- [16] Negara, E.S., 2017. Kajian terhadap tools dan framework social media analytics untuk pemanfaatan data social media dalam penelitian ilmu sosial. *Jurnal Teknologi Technoscientia*, pp.132-138.
- [17] Negara, E.S., Andryani, R. and Saksono, P.H., 2016. Analisis data twitter: Ekstraksi dan analisis data geospasial. *INKOM Journal*, 10(1), pp.27-36.