

## PREDIKSI PENYAKIT KARDIOVASKULAR BERBASIS ASOSIATION RULE

Taryadi<sup>1)</sup>, Era Yuniyanto<sup>2)</sup>  
STMIK Widya Pratama, Pekalongan <sup>12)</sup>  
tari\_ball@stmik-wp.ac.id<sup>1)</sup>, era.yuniyanto@gmail.com<sup>2)</sup>

### Abstrak

*Implementasi metode data mining dalam lingkup kesehatan merupakan bidang yang menarik dan sedang naik daun. Hal ini disebabkan semakin banyak tersedia data bidang kesehatan yang tersedia secara bebas sebagai bagian proses analisis dengan metode data mining. Saat ini data yang tersedia secara bebas tersebut telah banyak dianalisis dengan menggunakan metode data mining yang seperti klasifikasi dan klasterisasi, namun metode association rule belum banyak digunakan untuk menganalisis data tersebut. Metode data mining aturan asosiasi merupakan algoritma sederhana namun memiliki kekuatan yang memfokuskan pada hubungan tersembunyi antara atribut data dan melakukan validasi statistik. Hubungan yang terjadi antar atribut data dapat membantu untuk memahami penyakit dan penyebabnya dengan cara yang lebih baik, sehingga dapat membantu untuk mencegah terjadinya penyakit. Penelitian ini mengeksplorasi metode aturan asosiasi dengan menggunakan data penyakit jantung yang tersedia di repositori UCI. Penyakit kardiovaskular merupakan penyakit yang berhubungan dengan jantung dan sistem peredaran darah.*

**Kata kunci:** Data mining; healthcare; association rules; cardiovascular disease; heart disease

### 1. Pendahuluan

Teknologi informasi telah banyak mengubah proses perawatan kesehatan. Dalam domain perawatan kesehatan terdapat banyak informasi spesifik berhubungan dengan pasien yang dihasilkan, dipertukarkan, dan disimpan (Crockett, Eliason, 2017). Salah satu alasannya adalah semakin mudahnya media penyimpanan data. Catatan kesehatan secara elektronik tersedia di rumah sakit yang berisi informasi spesifik pasien dalam format yang mudah untuk dihitung, dikumpulkan dan disimpan sebagai data perawatan kesehatan secara klinis pasien. Repositori medis digunakan dengan berbagai cara dan merupakan peluang untuk penemuan baru.

Dengan menggunakan teknik pengembangan data, banyak informasi yang dapat dikumpulkan dari data pasien seperti demografi, pribadi, data tes yang telah dijalani pasien dan pendukung lainnya (Chaurasia, 2013). Data perawatan kesehatan merupakan potensi yang besar bagi industri untuk pengembangan dan digunakan bagi masyarakat. Informasi yang tersedia dianalisis untuk mengidentifikasi inefisiensi dan menyediakan layanan terbaik sehingga meningkatkan potensi perawatan kesehatan dan mengurangi biaya perawatan pasien.

Penyakit kanker, diabetes, ginjal, hipertensi, TBC, jantung, dan stroke berdampak besar pada kesehatan manusia. Penyakit tersebut merupakan penyakit jangka panjang yang memberikan dampak kronis dan umumnya progresif dan terdapat banyak faktor penyebab potensial terjadi penyakit tersebut. Penyakit kronis merupakan penyakit yang diderita oleh pasien lebih dari 6 bulan, dan saat ini merupakan penyebab utama kematian secara global. Organisasi Kesehatan Dunia (WHO) melaporkan bahwa sekitar 42% kematian penyakit tidak menular terjadi pada pasien dengan usia dibawah 70 tahun di seluruh dunia. Penyakit yang tidak menular yang menjadi perhatian adalah penyakit kardiovaskular, pernapasan kronis, kanker dan diabetes. Pasien yang memiliki salah satu dari penyakit ini perlu melakukan berbagai tes untuk menilai tingkat penyakit yang diderita. Terdapat beberapa faktor lain yang mempengaruhi penyakit tersebut seperti gaya hidup dan pola makan pasien. Jadi terdapat banyak atribut untuk menilai penyakit pasien. Dengan melakukan analisis data pasien secara khusus, akan menemukan hubungan antar atribut, sehingga dapat menjelaskan hubungan antar atribut dan informasi penyelamat hidup lainnya (Sudhakar, Manimekalai, 2017). Pengetahuan ini

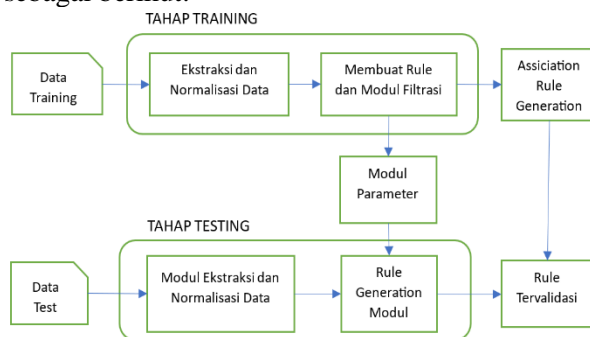
membantu untuk memahami sifat penyakit, hubungan antar atribut yang berbeda, bagian tubuh yang terkena penyakit dan terdampak dengan penyakit tersebut dan semua tes untuk menentukan tingkatan penyakit yang diderita pasien.

Dokter dapat mempelajari dan membuat model dari data yang tersedia dan menjadi alat untuk membantu memahami penyakit dengan cara yang lebih baik dan membantu pasien mengontrol dan mencegah komplikasi yang dapat timbul dimasa mendatang. Penelitian ini fokus pada penyakit kardiovaskular yaitu penyakit jantung, merupakan salah satu penyebab utama kematian di dunia. Jantung merupakan pusat seluruh tubuh dan faktor-faktor terkait dengannya tidak dipahami dengan baik oleh masyarakat. Pola antara berbagai faktor inilah yang ditemukan yang mengungkapkan beberapa relasi yang terjadi dan menarik (Jyoti, 2012).

Kontribusi dari penelitian ini adalah 1) mengadopsi modul awal untuk meningkatkan kemampuan menggali interior fitur visual dan fokus pada representasi fitur; 2) merancang modul untuk mengkompensasi hilangnya detail perseptual, target tersembunyi dan identifikasi obyek tersama dalam skala kecil; 3) mengevaluasi model dan membandingkannya dengan metode yang canggih.

## 2. Metode Penelitian

Tahapan penelitian ditunjukkan dengan diagram skematik desain terinci pada gambar 1. Penelitian ini menggunakan dua tahap pendekatan yaitu tahap pelatihan dan pengujian. Tahapan metodologi yang diusulkan dijelaskan sebagai berikut:



Gambar 1. Diagram skema pendekatan association rule mining

### 2.1. Ekstraksi dan Nominalisasi Data

Ekstraksi dan nominalisasi data merupakan fase awal pada tahap pelatihan. Dalam tahapan ini penyediaan data sesuai dengan format yang dibutuhkan perlu dilakukan. Pengabaian nilai yang hilang perlu dilakukan karena dapat mempengaruhi hasil yang salah dan menyesatkan. Tahapan ini merupakan tahap preprocessing data sebelum melakukan analisis. Data yang telah dibersihkan dengan membuang nilai yang hilang selanjutnya dilakukan proses perubahan format sesuai dengan kebutuhan analisis data.

Data yang akan dianalisis dalam bentuk numerik. Penggunaan algoritma Apriori sebagai salah satu model asosiasi rule model membutuhkan data dengan formal nominal, sehingga proses nominalisasi data harus dilakukan. Proses nominalisasi data dengan cara mengkategorikan atribut sesuai dengan rentang data. Misalnya atribut tekanan darah istirahat yang memiliki nilai kontinue dengan satuan mm Hg dapat dinominalisasi dengan mengatur pada rentang nilai tertentu.

### 2.2. Pembuatan aturan dan Filtrasi

Setelah tahapan preposing untuk menyiapkan data, tahapan selanjutnya adalah membuat aturan asosiasi dan filtrasi data. Tahapan untuk menyusun aturan asosiasi dapat dilakukan dengan menemukan sekumpulan atribut yang paling sering muncul. Terdapat dua tahapan untuk menemukan aturan asosiasi tersebut yaitu dengan cara : 1) menemukan itemset yang paling sering muncul sesuai dengan beberapa batasan minimum dan 2) menghitung frekuensi kemunculan dari item set tersebut. Aturan yang digunakan memiliki format  $X \Rightarrow Y$ , dengan menggunakan batasan kriteria tingkat kepercayaan minimal, dimana X dan Y saling eksklusif.

Algoritma Apriori (Agrawal, Srikant, 2014) merupakan algoritma data mining yang menggunakan aturan asosiasi untuk menarik kesimpulan dan sangat populer. Penelitian ini menggunakan algoritma Apriori untuk memprediksi penyakit jantung (Dominic, Gupta, Khare, 2015). Proses untuk menemukan frekuensi itemset dalam algoritma apriori dengan

menggunakan pembangkitan kandidat dan aturan. Algoritma Apriori akan menghasilkan aturan yang sangat besar sehingga akan sulit untuk menganalisis satu persatu. Proses penyaringan aturan yang relevan dilakukan untuk menyederhanakan aturan yang dihasilkan oleh algoritma ini (Agrawal, Imielinski, Swami, 2017).

Terdapat tiga jenis aturan asosiasi yang dihasilkan dengan menggunakan algoritma Apriori dengan mempertimbangkan keberadaan penyakit jantung, yaitu sebagai berikut:

1) Atribut => kelas :

Jenis aturan ini menunjukkan bahwa atribut atau daftar atribut menyebabkan penyakit jantung. Jenis aturan ini menjadi perhatian utama karena memberikan wawasan tentang faktor penyebab penyakit jantung.

2) Atribut => atribut :

Jenis aturan ini menunjukkan bahwa satu atribut atau daftar atribut menyebabkan atribut atau daftar atribut lainnya. Misalnya, lama di atas treadmill akan menunjukkan tekanan darah tinggi.

3) Kelas => atribut :

Jenis aturan ini menyiratkan bahwa beberapa atribut disebabkan oleh penyakit jantung. Aturan-aturan tersebut tidak penting untuk dianalisis karena fokus utama tulisan ini adalah untuk menemukan faktor penyebab penyakit jantung.

Fokus kajian yang dilakukan dalam makalah ini adalah membangkitkan aturan tipe Attribute => class, yaitu suatu bentuk klasifikasi seperti apa semua faktor yang berhubungan dengan tidak ditemukannya dan adanya penyakit jantung. Jenis aturan ini juga dikenal sebagai aturan asosiasi klasifikasi (Liu, Hsu, Ma, 2018)

Format aturan yang dihasilkan adalah :

$$\text{Antecedent} \Rightarrow \text{Consequent} \quad [support, confidence, lift]$$

Rumus untuk menghitung Rule-support, confidence dan lift menggunakan persamaan

berikut, dimana X dan Y merupakan sekumpulan atribut yang saling eksklusif:

$$\text{Rule } X \Rightarrow Y \quad (1)$$

$$\text{Support} = \frac{\text{frequency}(X, Y)}{N} \quad (2)$$

$$\text{Confidence} = \frac{\text{frequency}(X, Y)}{\text{freq}(X)} \quad (3)$$

$$\text{Lift} = \frac{\text{frequency}(X, Y)}{\text{freq}(X) \text{freq}(Y)} \quad (4)$$

Aturan  $X \Rightarrow Y$  memiliki arti setiap himpunan atribut X terdapat dalam dataset maka himpunan atribut Y dengan probabilitas sesuai dengan rumus (1). Dukungan aturan didefinisikan sebagai presentase baris dalam kumpulan data X dan Y dan ditunjukkan dengan rumus (2). Tingkat keyakinan aturan merepresentasikan dari persentase baris berisi X dan Y dengan jumlah baris yang berisi X dan dihitung dengan rumus (3). Lift aturan merupakan ukuran korelasi untuk menilai kemunculan suatu set atribut mengangkat kemunculan item atau sekumpulan item lainnya. Apabila nilai lift lebih besar dari 1 maka terdapat korelasi positif antara anteseden dan konsekuen.

Terdapat banyak aturan yang dihasilkan berdasarkan support. Akurasi membatasi jumlah aturan yang dihasilkan. Dengan menetapkan batasan support dan confidence untuk memfilter hitungan aturan dan mendapatkan hasil, tingkat lift digunakan untuk membatasi aturan untuk membatasi hanya aturan yang berkorelasi positif saja yang dihasilkan. Jadi support, confidence dan lift digunakan untuk memfilter jumlah aturan yang dihasilkan.

### 2.3. Pengujian dan validasi aturan

Tahap pengujian dilakukan dengan melakukan pengujian aturan yang dihasilkan pada tahap pelatihan, setelah melalui tahap preprocessing, ekstraksi data dan nominalisasi. Dataset pada tahap pengujian berbeda dengan dataset yang digunakan pada tahap pelatihan. parameter support dan confidence yang digunakan pada tahap pelatihan masih

dilanjutkan di tahap pengujian. Jika terdapat jumlah batasan kasus dalam data test mengikuti aturan asosiasi yang dihasilkan pada tahap pelatihan, maka aturan tersebut dinyatakan sebagai aturan yang valid. Rumus perhitungan akurasi untuk memverifikasi validitas aturan adalah sebagai berikut :

$$\text{Akurasi} = \frac{\text{Jumlah akurasi rule terklasifikasi}}{\text{Total jumlah rule yang dihasilkan}}$$

Akurasi merupakan rasio jumlah aturan yang diklasifikasikan pada tahap pengujian berbanding dengan total aturan yang dihasilkan pada tahap pelatihan.

### 3. Hasil dan Pembahasan

#### 3.1 Setup Uji Coba

Dataset yang digunakan untuk penambahan aturan asosiasi diambil dari repositori pembelajaran mesin populer UCI yang sering digunakan untuk berbagai aplikasi penambahan data (UCI Data Set).

Ada empat dataset yang tersedia untuk penyakit jantung yaitu dataset Cleaveland, dataset Hongaria, dataset Long Beach, dan dataset Swiss seperti yang ditunjukkan pada Tabel I. Dari empat ini, tiga digunakan dalam fase Pelatihan untuk Pembuatan Aturan dan satu digunakan dalam fase Pengujian.

Tabel 1. Data Statistik – Dataset Penyakit Jantung

Dataset	Atribut	Instan	Miss. Value
Long Beach	75	200	6%
Cleveland	75	282	0,1%
Hungarian	75	294	7%
Switzerland	75	123	11%

Tabel 2. Deskripsi Atribut Terpilih

Atribut	Keterangan
Age	Umur pasien
Sex	Jenis kelamin
Cp	Tipe cestpain
Trestbps	Resting blood pressure
Chol	Serum cholesterol
Fbs	Fating blood sugar > 120 mg/dl
restecg	Resting electrocardiofascular result
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina
Oldpeak	ST depression induced by exercise
Slope	The slope of the peak exercise
Ca	Number of major vessel
Thal	Thalasemia
Cigs	Number of cigarette smoked per day
Years	Number of years as a smoker
Dm	Histori of diabetes
Famhis	Family history of coronory heart disease
Dig	Digitalis used during exercise ECG
Prop	Beta blocker used during exercise ECG
Nitr	Nitrales used during exercise ECG
Pro	Calcium channel blocker used during exercise ECG
Diuretic	Diuretic used during excercise ECG
Type	Class label

Ada total 75 atribut termasuk atribut yang diprediksi, tetapi ada banyak nilai yang hilang. Mempertimbangkan hanya atribut yang tidak memiliki nilai yang hilang, jumlah total atribut berkurang menjadi 24. Dalam analisis yang dilakukan hanya aturan asosiasi yang dipertimbangkan yang mencakup atribut kelas yang terkait dengan diagnosis penyakit jantung. Atribut kelas memiliki lima nilai yang berbeda tergantung pada tingkat risiko penyakit kardiovaskular yaitu 0, 1, 2, 3, dan 4; di mana 0 menyiratkan tidak ada penyakit jantung dan peningkatan tingkat risiko penyakit kardiovaskular direpresentasikan sebagai 1, 2, 3, 4. Dalam analisis yang dilakukan dalam makalah ini atribut kelas dikategorikan sebagai 0 yaitu tidak adanya penyakit jantung atau 1 yaitu adanya jantung penyakit. Atribut kelas 1 mewakili semua tingkat risiko 1, 2, 3 dan 4 penyakit kardiovaskular.

Langkah pertama dalam Datamining adalah preprocessing data yang melibatkan penghapusan atribut dengan banyak data yang hilang dan pemilihan atribut yang relevan. Setelah banyak iterasi, dan berdasarkan rekomendasi ahli akhirnya dipilih 24 atribut yang relevan dari 75 atribut untuk analisis data. Sebagian besar atribut yang relevan adalah kategorikal dan beberapa

numerik seperti yang ditunjukkan pada Tabel 2 yang mencantumkan beberapa atribut relevan yang dipilih. Karena penambahan aturan asosiasi membutuhkan data dalam bentuk nominal, sehingga rentang dibentuk untuk atribut numerik berdasarkan sumber daya ahli, dan kategorisasi dibuat sesuai seperti yang ditunjukkan untuk beberapa atribut pada Tabel 3, misalnya jumlah rokok yang dihisap per hari dikategorikan rendah (1-3 batang rokok sehari), sedang (3-7 batang rokok sehari), tinggi (7-11 batang rokok sehari) dan sangat tinggi (>11 batang rokok sehari).

Algoritma apriori diterapkan pada dataset dengan 24 atribut (termasuk atribut kelas). Aturan dihasilkan dengan ambang dukungan minimum 10% dan ambang kepercayaan minimum 85%. Kemudian aturan disaring untuk mengambil aturan hanya tipe 1 (dan tipe 0), terutama karena tujuan analisis ini hanya untuk mempelajari faktor penyebab penyakit jantung atau tes mana yang penting dalam diagnosis penyakit jantung. Pemfilteran lebih lanjut dilakukan untuk menghapus aturan yang berlebihan dan subset. Aturan yang dihasilkan ditunjukkan pada gambar 2 (Aturan Tipe 0) dan gambar 3 (Aturan Tipe 1). Ambang batas peningkatan lebih besar dari 1, hanya kasus yang berkorelasi positif yang dipertimbangkan dalam kumpulan aturan akhir.

Tabel 3. Jangkauan Range Atribut

Atribut	Nilai
Trestbps (tekanan darah istirahat mm/Hg)	90-120: normal 120-140: tidak biasa 140-160: tinggi >160 : sangat tinggi
Cholesterol (chol)	110-120: normal 200-240: batas tinggi 240_250: tinggi >250 : sangat tinggi
Thalrest (detak jantung saat istirahat)	35-85: normal >85: tinggi
Cigs (jumlah rokok yang dihisap per hari)	0: kosong 1-3: rendah 3-7: sedang 7-11: tinggi >11: sangat tinggi

Dengan mempelajari empat set data dalam konteks penambahan aturan asosiasi, ditemukan bahwa Cleveland dan Hungaria paling baik

memiliki lebih banyak jumlah contoh seperti yang ditunjukkan pada Tabel 1. Oleh karena itu, kontribusi utama pada aturan asosiasi dilakukan oleh set data Cleveland bersama dengan set data lainnya Kumpulan data Swiss dan Long Beach. Kumpulan data Hungaria digunakan untuk menguji aturan tersebut untuk pemeriksaan validitas. Bagian pengkodean untuk pembuatan dan validasi aturan Asosiasi dilakukan dengan menggunakan bahasa pemrograman java. Jika anteseden aturan yang dihasilkan oleh dataset pelatihan sama dalam dataset pengujian yang menghasilkan konsekuensi yang sama dalam pengujian dataset untuk lebih dari 90% kasus, maka aturan tersebut dianggap sebagai aturan yang valid. Aturan akhir yang dihasilkan juga dianalisis dengan kasus aktual untuk memeriksa kelayakannya dan ditemukan bahwa ini sebenarnya adalah faktor penyebab prediksi penyakit jantung.

### 3.2 Hasil dan Analisis

Setelah menyelesaikan prosedur penambahan aturan asosiasi menggunakan Algoritma Apriori, dihasilkan 40 aturan. Beberapa yang paling menonjol diberikan di bawah ini. Rangkaian aturan pertama memiliki tipe 0 di RHS yang menyiratkan tidak adanya penyakit jantung, jika kondisi di LHS terpenuhi. Semua aturan memiliki kepercayaan lebih dari 85%.

Mempelajari aturan-aturan ini telah mengungkapkan hal-hal berikut:

- 1) Kolesterol tinggi dan nyeri dada asimtomatik menunjukkan adanya penyakit jantung.
- 2) Nyeri dada asimtomatik adalah salah satu atribut paling umum yang muncul sebagai anteseden dalam banyak aturan karena diketahui bahwa nyeri dada sering kali terkait dengan beberapa masalah terkait jantung.
- 3) Gula darah hampir tidak ada hubungannya dengan penyakit jantung, yaitu diabetes dan penyakit jantung berkorelasi negatif.
- 4) Kolesterol tinggi dan penyakit jantung berkorelasi positif.
- 5) Laki-laki memiliki risiko penyakit jantung lebih tinggi dibandingkan perempuan.
- 6) Laki-laki dengan penyakit jantung selalu dites positif pada tes fluoroskopi.

- 7) Orang yang lebih tua memiliki risiko penyakit jantung yang lebih tinggi daripada orang yang lebih muda.
- 8) Jika Thal (thalassemia adalah kelainan darah yang diwariskan di mana tubuh membuat bentuk hemoglobin yang tidak normal yang menyebabkan penghancuran sel darah merah yang berlebihan sehingga menyebabkan anemia) adalah normal dan angina yang diinduksi olahraga yaitu terkait dengan nyeri dada tidak ada, hal ini mencerminkan bahwa penyakit jantung tidak ada.
- 9) Jika Thal tidak normal (reversible\_defect) maka sangat terkait dengan adanya penyakit jantung seperti yang dapat dilihat pada aturan Tipe 1.
- 10) Jumlah pembuluh darah besar yang diwarnai dengan fluoroskopi, yang digunakan untuk melihat kerja pemompaan jantung yaitu ca adalah tes yang dilakukan untuk diagnosis penyakit jantung. Untuk jantung yang sehat ca=0 yang dapat dilihat terkait dengan aturan Tipe 0.
- 11) Oldpeak berhubungan dengan depresi ST yang diinduksi oleh latihan relatif terhadap istirahat. Depresi ST berhubungan dengan EKG. Jika nilai puncak lama kecil, maka risiko penyakit jantung rendah seperti yang terlihat pada aturan Tipe 0.

```
[ca=0, thal=normal, exercise_induced_angina=no] ==> [type=0]
[supp=0.3068592, conf=90.0, lift=1.6159916]

[ca=0, exercise_induced_angina=no, oldpeak=small] ==> [type=0]
[supp=0.35740072, conf=87.0, lift=1.5519525]

[thal=normal, exercise_induced_angina=no, fbs_more_than_120=no,
oldpeak=small] ==> [type=0] [supp=0.32129964, conf=87.0, lift=1.5593295]

[thal=normal, exercise_induced_angina=no, oldpeak=small] ==> [type=0]
[supp=0.36823106, conf=86.0, lift=1.5447786]

[ca=0, exercise_induced_angina=no] ==> [type=0]
[supp=0.38267148, conf=85.0, lift=1.5276794]

[ca=0, exercise_induced_angina=no, fbs_more_than_120=no, oldpeak=small]
==> [type=0] [supp=0.31407943, conf=85.0, lift=1.5242884]
```

Gambar 1. Instan dari asosiasi rule – tipe rule 0 yang dihasilkan dari dataset penyakit jantung

```
[thal=reversible_defect, cp=asymptomatic, exercise_induced_angina=yes]
==> [type=1]
[supp=0.15884477, conf=94.0, lift=2.125567]

[thal=reversible_defect, cp=asymptomatic, age=old] ==> [type=1]
[supp=0.15884477, conf=90.0, lift=2.038809]

[thal=reversible_defect, cp=asymptomatic] ==> [type=1]
[supp=0.22743683, conf=90.0, lift=2.0434427]

[thal=reversible_defect, cp=asymptomatic, fbs_more_than_120=no] ==>
[type=1]
[supp=0.19133574, conf=90.0, lift=2.0395944]

[thal=reversible_defect, exercise_induced_angina=yes] ==> [type=1]
[supp=0.1732852, conf=89.0, lift=2.018215]

[is_male=yes, cp=asymptomatic, exercise_induced_angina=yes] ==> [type=1]
[supp=0.18411553, conf=89.0, lift=2.0314927]

[is_male=yes, thal=reversible_defect, cp=asymptomatic,
fbs_more_than_120=no] ==> [type=1]
[supp=0.1696751, conf=89.0, lift=2.0134552]

[is_male=yes, thal=reversible_defect, cp=asymptomatic] ==> [type=1]
[supp=0.19494584, conf=89.0, lift=2.0099435]

[slope=flat, cp=asymptomatic, exercise_induced_angina=yes] ==> [type=1]
[supp=0.15162455, conf=89.0, lift=2.02895]

[is_male=yes, thal=reversible_defect, exercise_induced_angina=yes] ==>
[type=1]
[supp=0.15162455, conf=88.0, lift=1.9866803]

[slope=flat, is_male=yes, cp=asymptomatic] ==> [type=1]
[supp=0.16606498, conf=87.0, lift=1.9706155]

[cp=asymptomatic, exercise_induced_angina=yes, age=old] ==> [type=1]
[supp=0.15884477, conf=86.0, lift=1.9588557]

[cp=asymptomatic, exercise_induced_angina=yes] ==> [type=1]
[supp=0.2202166, conf=86.0, lift=1.9507043]

[is_male=yes, cp=asymptomatic, chol=high] ==> [type=1]
[supp=0.15884477, conf=86.0, lift=1.9588557]

[slope=flat, is_male=yes, cp=asymptomatic, fbs_more_than_120=no] ==>
[type=1]
[supp=0.15523465, conf=86.0, lift=1.9526229]

[thal=reversible_defect, restecg=left_ventricular_hypertrophy] ==> [type=1]
[supp=0.16245487, conf=85.0, lift=1.9277761]

[is_male=yes, cp=asymptomatic, restecg=left_ventricular_hypertrophy] ==>
[type=1]
[supp=0.1696751, conf=85.0, lift=1.9402385]

[slope=flat, thal=reversible_defect] ==> [type=1]
[supp=0.18772563, conf=85.0, lift=1.9355013]

[slope=flat, thal=reversible_defect, fbs_more_than_120=no] ==> [type=1]
[supp=0.1696751, conf=85.0, lift=1.9402385]

[slope=flat, is_male=yes, thal=reversible_defect] ==> [type=1]
[supp=0.15884477, conf=85.0, lift=1.9211854]
```

Gambar 2. Instan dari asosiasi rule – tipe yang dihasilkan dari dataset penyakit jantung

Pengujian aturan dilakukan pada kumpulan data Hungaria. Akurasi 90% dicapai dalam analisis ini.

#### 4. Kesimpulan dan Saran

Setelah mempelajari hubungan antara atribut yang berbeda, tidak ada penemuan mengejutkan yang dibuat, tetapi hasil yang diperoleh mendukung pengetahuan yang sudah ada. Sayangnya, tidak ada kumpulan data baru yang

besar dan tersedia untuk umum untuk dianalisis. Dengan data yang besar, penambangan aturan asosiasi dapat sangat membantu. Jika data tentang berbagai tes tersedia, maka hubungan antara tes dan penyakit jantung dapat dianalisis secara mendalam yang dapat membantu dokter untuk lebih memahami tes apa yang harus dilakukan dan kapan harus melakukan tes tersebut. Hal ini dapat menghemat sejumlah besar uang dan waktu yang dihabiskan oleh pasien.

Jika kumpulan data dengan sejumlah besar catatan tersedia, maka studi yang lebih terperinci dapat dilakukan yang akan lebih bermanfaat bagi dokter dan juga pasien.

#### DAFTAR PUSTAKA

- Crockett, D. Eliason, B. 2017. "What is datamining in healthcare?", Health Catalyst
- Fayyad U., Piatetsky-Shapiro G. and Smyth P. 1996. "From Data Mining to Knowledge Discovery: An Overview", in *Advances in Knowledge Discovery and Data Mining*. Cambridge, Mass.: MIT Press/AAAI Press.
- Jyoti, N.B.K. 2012, "A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic", *International Journal of Computer Applications* (0975 – 8887), September 2012, Volume 54–No.17
- Chaurasia, V. 2013, "Early Prediction of Heart Diseases Using Data Mining Techniques", *Carib J .SciTech*, 2013, Vol.1, 208-21
- Sudhakar, K, Manimekalai, M., 2017, "Study of Heart Disease Prediction using Data Mining", *International Journal of Advanced Research in Computer Science and Software Engineering*, January 2017, Volume 4, Issue 1
- Dominic, V., Gupta, D., Khare, S., 2015, "An Effective Performance Analysis of Machine Learning Techniques for CardioVascular Disease", *Applied Medical Informatics*. March 2015, Vol 36, No 1, pp: 23-32.
- Agrawal R., Imielinski T., Swami A. 2017, "Mining association rules between sets of items in large databases". In *ACM SIGMOD Conference*, pages 207–216.
- Agrawal, R. Srikant, R. 2014, "Fast Algorithms for Mining Association Rules", In *Proceedings of the 20th VLDB Conference*, Santiago, Chile.
- Liu, B. Hsu, W. Ma, Y. 2018, "Integrating classification and association rule mining". In *Knowledge Discovery and Data Mining Integrating*, pages 80–86.
- UCI Heart Disease Dataset. tersedia di: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>