



Tugas Mengupas Jurnal

Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees



Faris Saifullah
3124640034

Pendahuluan

- Penyakit jantung koroner (CHD) adalah penyebab kematian paling umum di Eropa dengan hampir dua juta kematian setiap tahun.
- aktor risiko CHD terbagi menjadi:
 - Non-modifiable: Usia, jenis kelamin, riwayat keluarga, faktor genetik.
 - Modifiable: Merokok, hipertensi, diabetes, kolesterol, dan lipoprotein.
- Mengembangkan sistem data mining berbasis pohon keputusan (C4.5) untuk menilai faktor risiko CHD dan mengurangi kejadian CHD.
- Mengidentifikasi aturan klasifikasi yang penting untuk mendukung pengelolaan pasien dan mengurangi biaya terapi.
- Berbagai penelitian telah dilakukan menggunakan metode seperti Framingham equation, algoritma pohon keputusan, dan teknologi lain seperti fuzzy modeling dan neural networks untuk menganalisis faktor risiko CHD.
- Beberapa studi sebelumnya menunjukkan keberhasilan model pohon keputusan dalam klasifikasi dan diagnosis CHD dengan akurasi tinggi.
- Studi ini fokus pada penggunaan data mining berbasis pohon keputusan untuk mengevaluasi faktor risiko CHD.
- Aturan klasifikasi yang diperoleh diharapkan membantu pengelolaan pasien secara lebih baik dan efisien.



Bahan dan Metode

A. Data Collection, Cleaning, and Coding

- Data diperoleh dari 1500 subjek CHD antara tahun 2003–2006 dan 2009 (300 subjek per tahun).
 - Subjek direkrut berdasarkan kriteria: riwayat MI, intervensi koroner perkutan (PCI), atau operasi bypass arteri koroner (CABG).
 - Proses melibatkan identifikasi dan penghapusan duplikasi, pengisian nilai yang hilang, dan pengkodean data sesuai Tabel I.
 - Data yang tidak lengkap, terutama hasil biokimia, menyebabkan pengurangan jumlah kasus (lihat Tabel II).
- Faktor Risiko Sebelum Kejadian
 - Tidak dapat diubah: Usia, jenis kelamin, riwayat keluarga (FH).
 - Dapat diubah: Riwayat merokok (SMBEF), hipertensi (HxHTN), diabetes (HxDM).
 - Faktor Risiko Setelah Kejadian
 - Dapat diubah: Merokok setelah kejadian (SMAFT), tekanan darah (SBP, DBP), kolesterol total (TC), HDL, LDL, trigliserida (TG), dan glukosa (GLU).



TABLE I
CODING OF RISK FACTORS

| | Risk Factor | Code 1 | Code 2 | Code 3 | Code 4 |
|--|--------------------|----------|------------|----------|----------|
| Risk Factors Before The Event: non modifiable | | | | | |
| 1 | AGE | 1: 34-50 | 2: 51-60 | 3: 61-70 | 4: 71-85 |
| 2 | SEX | M: MALE | F: FEMALE | | |
| 3 | FH | Y: YES | N: NO | | |
| Risk Factors Before The Event: modifiable | | | | | |
| 4 | SMBEF | Y: YES | N: NO | | |
| 5 | H _χ HTN | Y: YES | N: NO | | |
| 6 | H _χ DM | Y: YES | N: NO | | |
| Risk Factors After The Event: modifiable | | | | | |
| 1 | SMAFT | Y: YES | N: NO | | |
| 2 | SBP (mmHg) | L<100 | N: 100-130 | H≥130 | |
| 3 | DBP (mmHg) | L<60 | N: 60-85 | H≥85 | |
| 4 | TC (mg/dL) | N<190 | H≥190 | | |
| 5 | HDL (mg/dL) | | | | |
| | Women | L<50 | N: 50-60 | H≥60 | |
| | Men | L<40 | N: 40-60 | H≥60 | |
| 6 | LDL (mg/dL) | N<100 | H≥100 | | |
| 7 | TG (mg/dL) | N<150 | H≥150 | | |
| 8 | GLU (mg/dL) | N <110 | H≥110 | | |

L: low; N: normal, H: high, D: dangerous.

TABLE II
NO. OF CASES PER SET OF RULES/MODELS INVESTIGATED

| | Model | MI | PCI | CABG |
|-------|-------|-------------|-----------|-----------|
| | | N/Tr/Ev | N/Tr/Ev | N/Tr/Ev |
| Event | Yes | 378/75/75 | 72/36/36 | 86/43/43 |
| | No | 150/75/75 | 274/36/36 | 307/43/43 |
| | Total | 528/150/150 | 346/72/72 | 392/86/86 |

N: total no. of cases, Tr and Ev give the number of cases in training and evaluation, respectively.



Bahan dan Metode

B. Classification by Decision Trees

- Algoritma C4.5 menggunakan pendekatan divide-and-conquer untuk membangun pohon keputusan. Algoritma ini bekerja dari atas ke bawah, memilih atribut terbaik untuk memisahkan kelas pada setiap tahap, dan memproses sub masalah secara rekursif. Pemangkasan dilakukan dengan heuristik berbasis signifikansi statistik.
- Dalam penelitian ini, kriteria pemisahan berikut diselidiki, yang disajikan secara singkat: perolehan informasi, indeks gini, statistik chi-kuadrat rasio kemungkinan, rasio perolehan, dan ukuran jarak.

Algoritma untuk menghasilkan pohon keputusan:

1. Mulai dengan membuat simpul akar (root node).
2. Jika semua data memiliki kelas yang sama, simpul tersebut menjadi daun dengan label kelas tersebut.
3. Jika tidak ada atribut yang tersisa, simpul menjadi daun dengan label kelas mayoritas.
4. Pilih atribut terbaik menggunakan kriteria pemisahan untuk membagi data, lalu beri label pada simpul.
5. Hapus atribut yang digunakan dari daftar kandidat dan ulangi proses pada setiap subset data.
6. Jika subset kosong, tambahkan simpul daun dengan kelas mayoritas.
7. Ulangi hingga seluruh simpul selesai.

Hasil akhirnya adalah pohon keputusan yang mewakili data pelatihan.



Bahan dan Metode

B. Classification by Decision Trees

- Information Gain (IG), berdasarkan teori informasi Claude Shannon, digunakan untuk memilih atribut terbaik dengan InfoGain tertinggi. Meski efektif, IG memiliki kelemahan pada atribut dengan banyak nilai unik, sehingga gain ratio digunakan sebagai alternatif.

$$\text{InfoGain}(A) = \text{Info}(D) - \text{Info}_A(D) \quad (2.1)$$

where A is the attribute investigated.

$$\text{Info}(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (2.2)$$

where

p_i = probability(class i in dataset D);
 m = number of class values.

$$\text{Info}_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \text{Info}(D_j) \quad (2.3)$$

where

$|D_j|$ = number of observations with attribute value j in dataset D ;

$|D|$ = total number of observations in dataset D ;

D_j = sub dataset of D that contains attribute value j ;

v = all attribute values.



Bahan dan Metode

B. Classification by Decision Trees

- Indeks Gini (GI): Indeks Gini adalah kriteria ketidakmurnian yang membandingkan distribusi probabilitas atribut target.
- Statistik Rasio Kemungkinan Chi-Kuadrat (χ^2): Signifikansi statistik kriteria perolehan informasi dapat dinilai menggunakan statistik rasio kemungkinan chi-kuadrat.

$$\text{GiniIndex}(D) = \text{Gini}(D) - \sum_{j=1}^v p_j \times \text{Gini}(D_j) \quad (2.4)$$

$$\text{Gini}(D) = 1 - \sum_{i=1}^m p_i^2. \quad (2.5)$$

$$G^2(A, D) = 2 \times \ln(2) \times |D| \times \text{InfoGain}(A). \quad (2.6)$$



Bahan dan Metode

B. Classification by Decision Trees

- Gain Ratio (GR): Rasio gain menciptakan pohon keputusan yang tidak mengubah atribut dengan sejumlah besar nilai yang berbeda. Dengan kata lain, hal ini mengurangi jumlah informasi yang hilang.
- Pengukuran Jarak (DM): Mirip dengan GR, pengukuran jarak menormalkan kriteria ketidakmurnian (GI). Hal ini menunjukkan perlunya kembali ke keadaan normal dengan cara yang berbeda.

$$\text{GainRatio}(A) = \frac{\text{InfoGain}(A)}{\text{SplitInfo}_A(D)} \quad (2.7)$$

$$\text{SplitInfo}_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right). \quad (2.8)$$

$$\text{DM}(A) = \frac{\text{Gini}(D)}{- \sum_{j=1}^v \sum_{i=1}^m p_{ij} \times \log_2(p_{ij})}. \quad (2.9)$$



Bahan dan Metode

B. Classification by Decision Trees

- Tim kami menggunakan kriteria yang ditentukan di atas untuk membuat alat penambangan data yang mendukung metode pohon keputusan C4.5. Salah satu masalah praktis utama dengan pembelajaran pohon keputusan adalah overfitting. Oleh karena itu, pemangkasan digunakan untuk mencegah overfitting. Dengan menggunakan estimasi kesalahan Laplace, kami menerapkan proses pemangkasan bottom-up. Saat simpul daun terbentuk dan pohon keputusan dibangun, kesalahan Laplace.
- Saat algoritma kembali ke simpul akar, kesalahan simpul daun dihitung di simpul induk. Jika kesalahan simpul induk lebih kecil dari total kesalahan anak-anaknya, simpul induk dipangkas dan diganti dengan simpul daun berlabel kelas mayoritas. Jika tidak, pemangkasan dihentikan.

$$E(D) = \frac{N - n + m - 1}{N + m} \quad (2.10)$$

where

C = class value majority class in D ;

N = number of observations in D ;

n = number of observations has class value C .



Bahan dan Metode

C. Classification Models Investigated

- Model yang diselidiki meliputi:
 - MI: Subjek dengan infark miokard (MI) ditandai sebagai simptomatik, sisanya asimtomatik.
 - PCI: Subjek yang hanya menjalani PCI ditandai sebagai simptomatik, sisanya asimtomatik (MI dan PCI bersama dikecualikan).
 - CABG: Subjek yang hanya menjalani CABG ditandai sebagai simptomatik, sisanya asimtomatik (MI dan CABG bersama dikecualikan).
- Tiga subset pengujian dilakukan untuk setiap model:
 - Faktor risiko sebelum kejadian (B).
 - Faktor risiko setelah kejadian (A).
 - Faktor risiko sebelum dan setelah kejadian (B + A).
- Untuk setiap model dan kriteria pemisahan, dilakukan 20 kali pengujian dengan pengambilan sampel acak. Total 300 kali pengujian dilakukan (20×5 kriteria $\times 3$ dataset). Uji Wilcoxon juga dilakukan untuk mengevaluasi perbedaan signifikan antara kriteria pemisahan dan model (B, A, dan B + A) pada $p < 0,05$.



Bahan dan Metode

D. Performance Measures

- Untuk mengevaluasi kinerja, digunakan ukuran berikut:
 - Klasifikasi yang benar (%CC): Persentase catatan yang diklasifikasikan dengan benar, yaitu $(TP + TN)/N$.
 - Rasio positif sejati (%TP): Jumlah contoh positif yang diprediksi benar.
 - Rasio positif palsu (%FP): Jumlah contoh negatif yang diprediksi salah sebagai positif.
 - Rasio negatif sejati (%TN): Jumlah contoh negatif yang diprediksi benar.
 - Rasio negatif palsu (%FN): Jumlah contoh positif yang diprediksi salah sebagai negatif.
 - Sensitivitas: Fraksi contoh positif yang diprediksi benar, yaitu $TP/(TP + FN)$.
 - Spesifisitas: Fraksi contoh negatif yang diprediksi benar, yaitu $TN/(TN + FP)$.
 - Dukungan: Jumlah kasus di mana aturan berlaku atau diprediksi benar.
 - Keyakinan: Jumlah kasus di mana aturan berlaku (atau prediksinya benar), dinyatakan sebagai persentase dari semua kejadian di mana aturan tersebut berlaku. Misalnya, untuk aturan $X \rightarrow Z$, keyakinan adalah probabilitas bersyarat bahwa transaksi yang memiliki X juga mengandung Z.

$$\text{Support} = P(XZ) = \frac{\text{no of cases that satisfy } X \text{ and } Z}{|D|}.$$

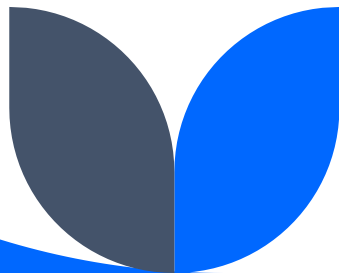
$$\text{Confidence} = P(Z|X) = \frac{P(XZ)}{P(X)}.$$



Bahan dan Metode

E. Calculation of the Risk

- Untuk setiap subjek, kami menghitung risiko peristiwa menggunakan persamaan Framingham. Subjek dibagi menjadi dua kategori: yang pernah mengalami peristiwa dan yang belum. Untuk setiap aturan yang diekstraksi, kami menemukan subjek yang sesuai dan menghitung rata-rata risiko peristiwa berdasarkan nilai risiko setiap subjek. Subjek diklasifikasikan sebagai risiko rendah ($<5\%$), menengah ($5-10\%$), atau tinggi ($>10\%$).



Hasil

- Tabel III menyajikan hasil klasifikasi dari tiga set model dengan lima kriteria pemisahan berbeda, menggunakan faktor risiko sebelum (B), setelah (A), dan sebelum serta setelah kejadian (B + A). Tabel ini menunjukkan median (Me), minimum (m), dan maksimum (M) untuk %CC, %TP, dan %FP, serta median untuk sensitivitas dan spesifisitas. Tabel IV mencantumkan tiga faktor risiko terpenting dari model pohon keputusan. Tabel V menyajikan aturan terpilih dari model tersebut, beserta risiko per aturan yang dihitung dengan persamaan Framingham.



Hasil

A. MI Models

- Tidak ditemukan perbedaan signifikan untuk kriteria pemisahan yang berbeda pada %CC menggunakan uji Wilcoxon dengan $p < 0,05$. Seperti yang terlihat dalam Tabel III, kinerja %CC pada model faktor risiko B, A, dan B + A berada di sekitar 60%, dengan model B + A menunjukkan kinerja terbaik, dengan median %CC antara 62% dan 63%. Model terbaik dicapai dengan kriteria pemisahan GI pada faktor risiko B + A, dengan %CC maksimum 66%. Faktor risiko terpenting, seperti yang tercantum dalam Tabel IV, adalah: untuk model B, usia, riwayat hipertensi, dan merokok sebelum kejadian; untuk model A, tekanan darah sistolik, merokok setelah kejadian, dan tekanan darah diastolik; serta untuk model B + A, usia, tekanan darah sistolik, merokok, dan riwayat hipertensi.
- Berdasarkan model pohon keputusan, contoh aturan dapat diekstraksi. Misalnya, seperti yang diberikan pada Tabel IV:
 1. Persentase subjek berusia 51–60 tahun dengan riwayat hipertensi yang bukan perokok dan mengalami kejadian hampir sama dengan yang merupakan perokok dan tidak mengalami episode. Untuk model MI, terdapat 0/0 (0/0%), 28/7 (5,3/1,3%), dan 330/163 (62,5/30,9%) subjek dengan kejadian ya/tidak, masing-masing dengan risiko rendah, menengah, dan tinggi. Risiko kejadian rata-rata per aturan berkisar antara 11,8% hingga 15,0%, dengan semua aturan diklasifikasikan sebagai risiko tinggi (lihat Tabel IV). Tidak ada perbedaan antara risiko kejadian aturan untuk kejadian MI terjadi atau tidak.



Hasil

TABLE III
CLASSIFICATION RESULTS OF THE THREE SET OF MODELS INVESTIGATED FOR THE FIVE DIFFERENT SPLITTING CRITERIA USING RISK FACTORS BEFORE THE EVENT (B), AFTER THE EVENT (A), AND BEFORE AND AFTER (B+A)

| | %CC | | | %TP | | | %FP | | | Sensitivity | | | Specificity | | |
|------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-------------|----|-----|-------------|----|-----|
| | B | A | B+A | B | A | B+A | B | A | B+A | B | A | B+A | B | A | B+A |
| | Me(m,M) | Me(m,M) | Me(m,M) | Me(m,M) | Me(m,M) | Me(m,M) | Me(m,M) | Me(m,M) | Me(m,M) | Me | Me | Me | Me | Me | Me |
| MI | | | | | | | | | | | | | | | |
| IG | 58(57,64) | 61(60,63) | 62(61,65) | 64(60,76) | 68(61,73) | 67(53,68) | 48(44,55) | 45(41,49) | 37(25,47) | 58 | 60 | 63 | 60 | 64 | 63 |
| GI | 61(59,63) | 61(59,63) | 63(61,66) | 67(55,71) | 59(55,71) | 63(57,76) | 47(41,59) | 36(33,48) | 39(25,51) | 59 | 60 | 62 | 61 | 62 | 64 |
| X2 | 58(57,60) | 61(59,63) | 63(62,65) | 65(63,73) | 63(59,76) | 64(59,72) | 49(47,53) | 39(35,59) | 36(35,47) | 57 | 62 | 64 | 59 | 61 | 64 |
| GR | 60(58,61) | 59(59,59) | 62(61,64) | 65(53,72) | 59(55,67) | 65(53,67) | 45(37,53) | 41(36,49) | 41(38,45) | 59 | 59 | 62 | 61 | 59 | 62 |
| DM | 60(58,62) | 59(58,62) | 63(61,65) | 71(57,67) | 61(57,69) | 65(57,71) | 47(39,54) | 43(40,45) | 40(27,45) | 59 | 59 | 65 | 63 | 59 | 64 |
| PCI | | | | | | | | | | | | | | | |
| IG | 63(61,65) | 67(64,75) | 67(65,70) | 64(53,72) | 72(67,78) | 58(56,64) | 36(31,42) | 39(28,50) | 22(22,31) | 63 | 65 | 71 | 63 | 69 | 65 |
| GI | 61(61,64) | 67(65,68) | 67(63,70) | 67(50,86) | 69(50,75) | 67(56,69) | 39(28,64) | 42(14,50) | 31(22,42) | 63 | 64 | 69 | 65 | 64 | 64 |
| X2 | 63(60,64) | 65(63,72) | 65(63,65) | 69(56,69) | 72(58,78) | 72(58,78) | 36(33,44) | 36(33,42) | 42(28,53) | 61 | 64 | 63 | 65 | 65 | 68 |
| GR | 63(61,70) | 64(64,65) | 65(64,67) | 67(56,82) | 67(53,83) | 72(53,72) | 44(31,50) | 39(25,56) | 39(22,44) | 65 | 63 | 65 | 63 | 65 | 67 |
| DM | 64(63,65) | 65(61,71) | 65(64,68) | 69(64,78) | 72(67,78) | 69(64,75) | 42(33,47) | 42(36,56) | 39(33,47) | 63 | 62 | 64 | 66 | 67 | 67 |
| CABG | | | | | | | | | | | | | | | |
| IG | 69(67,73) | 66(63,69) | 70(70,71) | 70(63,77) | 74(65,79) | 65(63,65) | 35(23,40) | 42(33,47) | 23(11,26) | 67 | 67 | 73 | 70 | 68 | 68 |
| GI | 69(69,71) | 63(61,65) | 69(67,71) | 67(58,74) | 67(56,72) | 74(72,74) | 28(21,35) | 42(30,42) | 37(33,40) | 70 | 63 | 67 | 68 | 64 | 70 |
| X2 | 69(67,73) | 63(61,65) | 69(67,72) | 72(63,81) | 72(63,79) | 74(72,77) | 33(21,44) | 47(42,58) | 37(30,42) | 67 | 61 | 67 | 69 | 66 | 71 |
| GR | 69(66,71) | 63(61,66) | 69(69,75) | 67(65,74) | 70(61,74) | 74(65,77) | 35(26,37) | 44(28,49) | 30(26,40) | 67 | 62 | 69 | 68 | 65 | 71 |
| DM | 71(70,72) | 61(59,67) | 69(69,71) | 67(63,72) | 77(58,81) | 70(58,74) | 28(19,30) | 49(40,58) | 33(21,35) | 73 | 59 | 70 | 71 | 67 | 70 |

The median (Me), (minimum (*m*), and maximum (*M*) for 20 runs are given for %CC, %TP, and %FP, whereas for sensitivity and specificity only the median values are given.



Hasil

TABLE IV
THREE MOST IMPORTANT RISK FACTORS OF THE THREE SET OF MODELS INVESTIGATED GIVEN IN TABLE III FOR THE FIVE DIFFERENT SPLITTING CRITERIA USING RISK FACTORS BEFORE THE EVENT (B), AFTER THE EVENT (A), AND BEFORE AND AFTER (B + A)

| | B | | | A | | | B+A | | |
|------|-----|-------|-------|-------|-------|-------|------|-------|-------|
| MI | | | | | | | | | |
| IG | AGE | SMBEF | HxHTN | SBP | SMAFT | DBP | AGE | SMAFT | SBP |
| GI | AGE | HxHTN | SMBEF | SBP | SMAFT | DBP | AGE | SBP | SMBEF |
| X2 | AGE | HxHTN | SMBEF | SMAFT | SBP | DBP | AGE | DBP | HxHTN |
| GR | AGE | HxHTN | SMBEF | SBP | SMAFT | DBP | SBP | SMAFT | HxHTN |
| DM | AGE | HxHTN | SMBEF | SBP | DBP | SMAFT | AGE | SBP | SMBEF |
| PCI | | | | | | | | | |
| IG | FH | AGE | HxDM | DBP | LDL | SMAFT | HxDM | DBP | FH |
| GI | AGE | HxHTN | FH | DBP | LDL | SMAFT | DBP | FH | HxHTN |
| X2 | FH | HxHTN | HxDM | DBP | LDL | SMAFT | DBP | HxHTN | AGE |
| GR | FH | HxHTN | HxDM | DBP | SMAFT | LDL | HxDM | FH | DBP |
| DM | FH | HxHTN | HxDM | DBP | LDL | SMAFT | FH | DBP | HxDM |
| CABG | | | | | | | | | |
| IG | AGE | HxHTN | SMBEF | SMAFT | SBP | DBP | AGE | SMBEF | HxDM |
| GI | AGE | HxDM | SMBEF | SMAFT | SBP | DBP | AGE | SMBEF | HxDM |
| X2 | AGE | SMBEF | HxDM | SMAFT | SBP | DBP | AGE | SMBEF | SMAFT |
| GR | AGE | HxDM | SMBEF | SMAFT | SBP | DBP | AGE | SMAFT | HxDM |
| DM | AGE | HxDM | SMBEF | SMAFT | DBP | SBP | AGE | SMAFT | HxDM |



Hasil

B. PCI Models

- Untuk model PCI, kinerja sedikit lebih baik dibandingkan dengan model MI, dengan kinerja terbaik pada model A dan B + A, di mana median %CC berkisar antara 65% hingga 67%. Kinerja serupa diperoleh untuk semua kriteria pemisahan tanpa perbedaan signifikan. Faktor risiko utama untuk model B meliputi usia, riwayat keluarga, riwayat hipertensi, dan diabetes; untuk model A, tekanan darah diastolik, LDL, dan merokok setelah kejadian; dan untuk model B + A, riwayat diabetes, tekanan darah diastolik, riwayat keluarga, hipertensi, dan usia. Aturan yang diberikan dalam Tabel IV menunjukkan bahwa jumlah kejadian PCI meningkat seiring bertambahnya usia pada subjek diabetes (dukungan meningkat dari 2% menjadi 20%). Untuk model PCI, terdapat 0/0 (0/0%), 20/15 (3,8/2,8%), dan 193/300 (36,6/56,8%) subjek dengan risiko rendah, menengah, dan tinggi. Risiko kejadian rata-rata per aturan berkisar antara 11,7% hingga 13,9%, dengan semua aturan diklasifikasikan sebagai risiko tinggi. Tidak ada perbedaan antara risiko kejadian PCI yang terjadi atau tidak.



Hasil

TABLE V
SELECTED RULES FROM MODELS GIVEN IN TABLE III (BASED ON THE CODING
OF THE RISK FACTORS GIVEN IN TABLE II)

| | SEX | | AGE | | | | FH | | SM | | HxHTN | | HxDM | | CLASS | | SUP % | CONF % | EVENT RISK | | |
|--------------------------------------|-----|---|-----|---|---|---|----|---|------------|---|-------|---|------|---|-------|---|-------|--------|------------|------|--|
| | M | F | 1 | 2 | 3 | 4 | Y | N | Y | N | Y | N | Y | N | Y | N | | | Y | N | |
| non Modifiable | | | | | | | | | Modifiable | | | | | | | | | | | | |
| Risk factors before the event (MI) | | | | | | | | | | | | | | | | | | | | | |
| 1.1 | | | + | | | | | | | | | | | | + | | 19 | 79 | 11,8 | 12,6 | |
| 1.2 | | | | + | | | | | | | | + | | | + | | 22 | 76 | 12,4 | 11,4 | |
| 1.3 | | | | + | | | | | | + | + | | | | + | | 10 | 67 | 12,6 | 12,4 | |
| 1.4 | | | | + | | | | | + | | + | | | | | + | 17 | 68 | 13,5 | 13,2 | |
| 1.5 | | | | | + | | | + | | | | + | | | + | | 20 | 63 | 12,7 | 12,9 | |
| 1.6 | | | | | + | | | + | | | + | | | | + | | 23 | 59 | 12,8 | 13,3 | |
| 1.7 | | + | | | | + | | | | | | | | | + | | 11 | 69 | 12,5 | 13,2 | |
| 1.8 | + | | | | | + | | + | | | | | | | + | | 24 | 61 | 12,1 | 13,4 | |
| 1.9 | + | | | | | + | + | | | + | | | | | + | | 7 | 64 | 12,6 | 12,9 | |
| 1.10 | + | | | | | + | + | | + | | | | | | + | | 10 | 67 | 15,0 | 14,3 | |
| Risk factors before the event (PCI) | | | | | | | | | | | | | | | | | | | | | |
| 2.1 | | | | | | | | | + | + | + | + | + | + | + | | 29 | 71 | 11,7 | 12,1 | |
| 2.2 | | | | | | | | | + | | | + | | + | + | + | 35 | 64 | 12,3 | 12,4 | |
| 2.3 | | | | | | | | | | | + | | | + | + | | 72 | 65 | 12,8 | 13,0 | |
| 2.4 | | + | | | | | | | | | | | + | | + | | 13 | 67 | 13,1 | 12,9 | |
| 2.5 | + | | + | | | | | | | | | | + | | + | | 2 | 100 | 13,1 | 12,0 | |
| 2.6 | + | | | + | | | | | | | | | + | | + | | 10 | 86 | 13,1 | 13,8 | |
| 2.7 | + | | | | + | | | | | | | | + | | + | | 21 | 67 | 13,1 | 13,3 | |
| 2.8 | + | | | | | + | | | | | | | + | | + | | 20 | 93 | 13,3 | 13,9 | |
| Risk factors before the event (CABG) | | | | | | | | | | | | | | | | | | | | | |
| 3.1 | | | + | | | | | | | | | | | | + | | 20 | 94 | 11,5 | 11,9 | |
| 3.2 | | | | + | | | | | | | | | | + | + | | 34 | 79 | 12,7 | 12,4 | |
| 3.3 | | | | + | | | | | | | | | + | | + | | 14 | 67 | 13,8 | 13,2 | |
| 3.4 | | | | | + | | | | | | + | | + | + | + | | 16 | 64 | 13,0 | 12,5 | |
| 3.5 | | | | | + | | | + | | | + | | + | + | + | | 16 | 57 | 12,7 | 12,8 | |
| 3.6 | | | | | + | | + | | | | + | | + | | + | | 19 | 69 | 13,3 | 12,7 | |
| 3.7 | | | | | + | | | | | | | | + | | + | | 28 | 71 | 13,0 | 13,3 | |
| 3.8 | | | | | | + | | | | | | | | | + | | 53 | 70 | 13,4 | 12,9 | |

Hasil

C. CABG Models

- Untuk model PCI, kinerja sedikit lebih baik dibandingkan dengan model MI, dengan kinerja terbaik pada model A dan B + A, di mana median %CC berkisar antara 65% hingga 67%. Kinerja serupa diperoleh untuk semua kriteria pemisahan tanpa perbedaan signifikan. Faktor risiko utama untuk model B meliputi usia, riwayat keluarga, riwayat hipertensi, dan diabetes; untuk model A, tekanan darah diastolik, LDL, dan merokok setelah kejadian; dan untuk model B + A, riwayat diabetes, tekanan darah diastolik, riwayat keluarga, hipertensi, dan usia. Aturan yang diberikan dalam Tabel IV menunjukkan bahwa jumlah kejadian PCI meningkat seiring bertambahnya usia pada subjek diabetes (dukungan meningkat dari 2% menjadi 20%). Untuk model PCI, terdapat 0/0 (0/0%), 20/15 (3,8/2,8%), dan 193/300 (36,6/56,8%) subjek dengan risiko rendah, menengah, dan tinggi. Risiko kejadian rata-rata per aturan berkisar antara 11,7% hingga 13,9%, dengan semua aturan diklasifikasikan sebagai risiko tinggi. Tidak ada perbedaan antara risiko kejadian PCI yang terjadi atau tidak.



Diskusi

- Penelitian ini menyelidiki MI, PCI, dan CABG dengan mengembangkan tiga model klasifikasi berbasis pohon keputusan, menghasilkan tingkat akurasi tertinggi masing-masing 66%, 75%, dan 75%. Faktor risiko utama yang ditemukan adalah jenis kelamin, usia, merokok, tekanan darah, dan kolesterol, dengan tiga faktor terakhir dapat dimodifikasi untuk mengurangi risiko PJK melalui pengendalian yang tepat. Merokok terbukti secara signifikan meningkatkan risiko PJK.
- Penelitian ini sejalan dengan temuan studi EUROASPIRE (I, II, III), yang melibatkan berbagai populasi Eropa dan menyoroti faktor risiko tambahan seperti obesitas. Hasil umum menunjukkan rendahnya kepatuhan pasien terhadap saran medis. Dibandingkan EUROASPIRE, temuan kami menunjukkan: 14% pasien masih merokok (16% di EUROASPIRE), 22% memiliki tekanan darah tinggi (26%), 34% memiliki kolesterol total tinggi (31%), dan 45% memiliki lipoprotein densitas rendah (31%).
- Survei EUROASPIRE mengidentifikasi merokok, tekanan darah, dan kolesterol sebagai faktor risiko utama, dengan variasi prevalensi yang signifikan di antara 15 negara serta penggunaan terapi obat kardioprotektif. Survei ini menyimpulkan bahwa masih ada potensi besar untuk meningkatkan standar perawatan pencegahan di seluruh Eropa guna mengurangi risiko penyakit berulang dan kematian pada pasien PJK.
- Studi ini mengungkapkan pengamatan tambahan terkait faktor risiko yang tidak dapat dimodifikasi dibandingkan dengan survei EUROASPIRE, yaitu: 14% subjek adalah perempuan (24,7% di EUROASPIRE); 9% berusia ≤ 50 tahun (23,1%); 28% berusia 51–60 tahun (33,8%); 39% berusia 61–70 tahun (43,1%); dan 24% berusia 71–84 tahun.



Diskusi

- Penelitian ini menunjukkan bahwa merokok meningkatkan risiko kejadian koroner berulang, sementara faktor risiko seperti kolesterol dan kelebihan berat badan juga relevan. Menggunakan faktor risiko tradisional, seperti usia, jenis kelamin, kolesterol, tekanan darah, diabetes, dan merokok, untuk memprediksi PJK dengan fungsi Framingham, namun menyimpulkan bahwa faktor risiko ini memiliki dampak yang berbeda dan memerlukan fungsi risiko khusus populasi. Sebagian besar subjek dalam penelitian ini tergolong berisiko tinggi, sehingga model yang diusulkan membutuhkan validasi pada kelompok yang lebih heterogen.
- Penambahan data berbasis pohon keputusan digunakan untuk mengekstrak aturan risiko PJK, dengan kinerja model serupa meskipun menggunakan berbagai kriteria pemisahan. Studi ini menegaskan kegunaan model ini dalam mengelompokkan faktor risiko menjadi tinggi dan rendah, serta potensinya untuk membantu pengambilan keputusan terapi, seperti angioplasti atau pembedahan. Namun, diperlukan investigasi lebih lanjut dengan dataset yang lebih besar dan algoritma yang lebih beragam untuk meningkatkan keandalan model dan mengurangi morbiditas serta mortalitas PJK.



Sumber Jurnal

Referensi : <https://ieeexplore.ieee.org/document/5378501/>

