

KLASIFIKASI POTENSI MENDERITA PENYAKIT JANTUNG KORONER DENGAN BACKPROPAGATION NEURAL NETWORK

Henri ¹⁾, Chairisni Lubis ²⁾

¹⁾²⁾ Teknik Informatika, FTI, Universitas Tarumanaraga
Jl. Letjen S Parman no 1, Jakarta 11440 Indonesia
email : henri.535180074@stu.untar.ac.id¹⁾, chairisnil@fti.untar.ac.id²⁾

ABSTRACT

Maksimal 200 kata berbahasa Inggris dicetak miring dengan Times New Roman 10 point. Abstrak berisi maksud dan tujuan penelitian, metode penelitian yang digunakan serta hasil yang dicapai.

Membandingkan algoritma A dan algoritma B dalam menyelesaikan sistem persamaan linier berdasarkan kecepatan proses, ketelitian hasil serta besarnya ruang memori yang digunakan. Hasil percobaan menunjukkan algoritma B lebih cepat dari algoritma A namun algoritma A menghasilkan ketelitian yang lebih baik serta ruang memori yang kecil dibandingkan algoritma B

Key words

Back Propagation Neural Network, Penyakit Jantung Koroner, SMOTE

1. Pendahuluan

Pada masa pandemi, masyarakat mengurangi aktivitas di luar rumah dan memperbanyak melakukan aktivitas dalam ruangan, dalam masa pandemi diberlakukan pembatasan sosial yang secara langsung maupun tidak langsung mengurangi tingkat kegiatan masyarakat secara signifikan. Turunnya aktivitas fisik juga berkaitan dengan menurunnya metabolisme tubuh yang dapat meningkatkan resiko penyakit, diantaranya penyakit jantung koroner.

Penyakit jantung sendiri merupakan penyebab kematian nomor satu di dunia. Organisasi Kesehatan Dunia (WHO) memperkirakan 17.9 juta kematian yang merupakan 32% dari total kematian di dunia pada tahun 2019. Sebesar 75% kematian akibat penyakit kardiovaskular terjadi di negara berpenghasilan rendah dan negara berpenghasilan menengah [1].

Seiring perkembangan bidang teknologi informasi. Dengan menerapkan Artificial neural Network dalam melakukan klasifikasi, diharapkan dapat memberikan masyarakat luas akses untuk melakukan pra-diagnosa penyakit jantung koroner tanpa harus melalui tenaga medis yang kemungkinan tidak semua kalangan masyarakat memiliki kemampuan untuk konsultasi ke dokter spesialis jantung. penerapan Artificial Intelligence

dapat berfokus untuk melakukan pra-diagnosa tanpa menggantikan peran seorang dokter spesialis jantung.

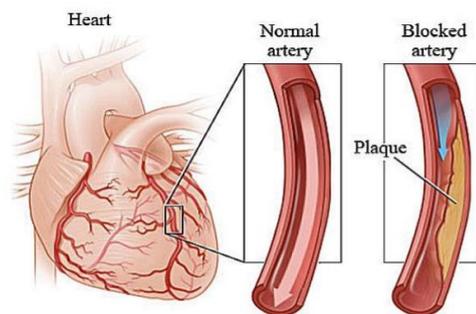
Metode penelitian belum tau gimana?

Penelitian sebelumnya belum tau gimana?. Pada penelitian kali ini akan menerapkan Back Propagation Neural Network dan melakukan oversampling dengan metode Synthetic Minority Over-Sampling Technique (SMOTE). teknik ini akan membuat dataset sintesis dengan menerapkan algoritma K-Nearest Neighbor untuk menentukan letak node sintesis yang akan dibuat.

2. Landasan Teoretik

2.1 Penyakit Jantung Koroner

Penyakit jantung koroner sering dikaitkan dengan aterosklerosis. Aterosklerosis digambarkan sebagai penumpukan bahan lemak dan kolesterol yang berkonsistensi lunak dan atau kalsium yang mengeras di sepanjang dinding arteri. Bentuk inilah yang dikenal dengan plak aterosklerosis. Plak ini akan menyumbat sebagian atau seluruh lumen arteri [2].



Gambar 1 Sumbatan Plak Pada Pembuluh Darah

Proses aterosklerosis ditandai dengan terbentuknya fatty streak. Fatty streak akan berkembang sejalan dengan usia. Progresivitas fatty streak untuk berkembang menjadi aterosklerosis, sangat dipengaruhi oleh ada tidaknya faktor-faktor risiko yang menyertainya. Semakin banyak faktor risiko yang mendasarinya akan menyebabkan semakin beratnya proses aterosklerosis [3]. Faktor

risiko terdiri dari faktor risiko yang dapat dimodifikasi dan tidak dapat dimodifikasi. Yang tidak dapat dimodifikasi antara lain: usia, jenis kelamin dan Riwayat penyakit. Sedangkan faktor risiko penyakit jantung koroner yang dapat dimodifikasi antara lain hipertensi, merokok, kurangnya aktivitas fisik, obesitas, diabetes melitus dan makanan yang dikonsumsi. Faktor risiko yang penyakit jantung koroner bersifat kumulatif, yang berarti semakin banyak faktor risiko yang dimiliki, maka risiko untuk menderita penyakit kardiovaskular semakin. Adapun Faktor risiko yang bersifat major-independennya terdiri dari hipertensi, diabetes melitus, kebiasaan merokok dan kadar kolesterol tinggi.

2.2 Synthetic Minority Over-Sampling Technique (SMOTE)

Data Penyakit merupakan data yang cenderung tidak seimbang. Dari suatu dataset akan terdapat lebih banyak kasus negatif daripada kasus positif. Hal ini akan membuat dataset yang tidak seimbang dan menghambat performa pelatihan model Back propagation Neural Network hingga menyebabkan ketidak akuratan dari hasil uji.

Smote akan melakukan data sampling pada kelas minority data dengan menggunakan K-Nearest Neighbor (KNN), KNN merupakan metode klasifikasi klasik yang paling sederhana. Metode ini melakukan klasifikasi terhadap object berdasarkan jarak antara object tersebut dengan object lain. Metode ini akan membuat kelas baru dengan mengambil k tetangga terdekat pertama kemudian dihitung jumlah data yang mengikuti kelas yang ada dari k tetangga tersebut [4].

KNN digunakan untuk menentukan dimana letak node baru yang akan dibuat. KNN menghitung jarak antar node dengan menggunakan fungsi Euclidean. Fungsi Euclidean ini dapat dilihat pada persamaan (1).

$$\sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

Setelah didapatkan sejumlah node terdekat pada node minoritas terpilih dari fungsi euclidean pada persamaan (1). Maka akan dipilih satu node secara acak dari sekelompok node terdekat yang terpilih sebelumnya dan node baru akan dibuat diantara kedua node tersebut. Node akan secara acak diantara kedua node dengan r (2).

$$r = \text{random}(0.0, 1.0) \quad (2)$$

Setelah didapatkan nilai acak yang akan menentukan letak node baru, nilai tiap-tiap dimensi node baru akan dibuat dengan fungsi (3) dimana n mewakili index tiap-tiap dimensi.

$$p_{n_new} = p_n + ((q_n - p_n) * r) \quad (3)$$

Dengan menggunakan cara diatas tidak bekerja maksimal pada data nominal, data nominal adalah data yang diwakili dengan notasi 0 dan 1. Untuk feature yang memiliki nilai nominal, perhitungan jarak pada dimensi

dilakukan dengan. Std_{med} yang merupakan standart deviasi dari setiap dimensi yang memiliki feature continuous Ir yang merupakan rasio kelas minoritas terhadap semua data.

Berikutnya akan dilakukan encoding pada tiap nilai nominal pada tiap-tiap dimensi yang memiliki nilai nominal. Encoding ini dilakukan dengan fungsi (4) dan (5).

$$e' = e * ir \quad (4)$$

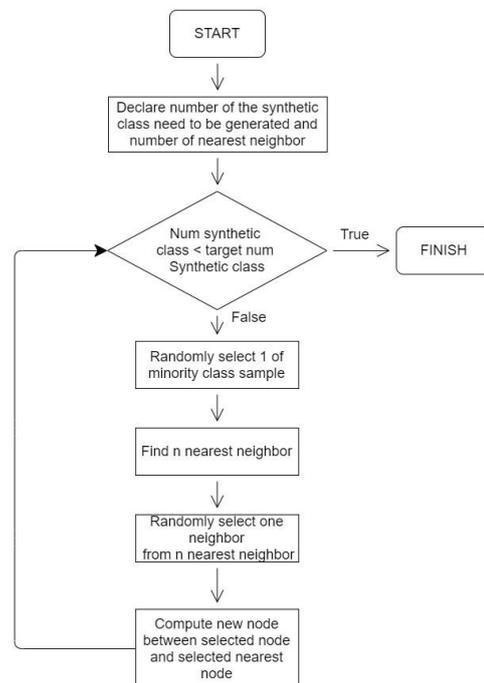
$$\text{encoded label value} = \frac{(o - e')}{e'} * Std_{med} \quad (5)$$

Dimana e adalah frekuensi nilai nominal tersebut muncul pada seluruh data, e' adalah hasil dari frekuensi nilai nominal dikalikan dengan imbalance ratio (ir). o adalah frekuensi nilai nominal tersebut muncul pada data minoritas.

Setelah mendapatkan semua nilai encoded label value in dimension pada tiap value nominal ditiap dimensi, perhitungan dimensi bernilai nominal pada perhitungan jarak euclidean akan berubah menjadi seperti fungsi (6)

$$\sqrt{.. + (\text{encoded label '0'} - \text{encoded label '1'})^2} \quad (6)$$

Selain perhitungan jarak, proses pembuatan nilai pada dimensi yang bernilai nominal juga berubah, nilai yang akan mengisi dimensi yang bernilai nominal didapatkan dengan melakukan random yang akan menghasilkan nilai 0 atau 1.



Gambar 2 Flowchart SMOTE

2.2 Back Propagation Neural Network

Menurut Winston, 1992, Artificial Intelligence adalah sebuah bidang ilmu yang memungkinkan komputasi digital untuk menerima, memahami dan bertindak [5] Secara umum, Artificial Intelligence adalah suatu bidang untuk meniru cara kerja syaraf manusia dan diterapkan dalam suatu komputasi digital.

BPNN merupakan algoritma pelatihan terbimbing yang mempunyai banyak lapisan. BPNN menggunakan error output untuk mengubah nilai bobot-bobotnya dalam arah mundur(backward). Untuk mendapatkan error ini, tahap perambatan maju (forward propagation) harus dikerjakan terlebih dahulu.

Dalam ketiga fase dalam pelatihan BPNN, yaitu fase maju (feed forward), fase mundur (back propagation) dan fase modifikasi bobot. Dalam fase feed forward, pola masukan dihitung maju dimulai dari lapisan input hingga lapisan output. Dalam fase back propagation, tiap-tiap unit output menerima target pola yang terhubung dengan pola input untuk dihitung nilai kesalahan. Kesalahan tersebut akan dipropagasikan mundur. Sedangkan fase modifikasi bobot bertujuan untuk menurunkan kesalahan yang terjadi. Ketiga fase tersebut akan diulang terus menerus hingga kondisi penghentian tercapai. Algoritma BPNN adalah sebagai berikut

1. Inisialisasi Bobot dengan bilangan acak bernilai kecil
2. Selama kondisi berhenti belum terpenuhi, ulangi Langkah 3 sampai 8
3. Tiap unit masukan ($x_i, i = 1, \dots, n$) menerima masukan x_i dan diteruskan ke hidden layer.
4. Tiap Unit tersembunyi ($z_j, j = 1, \dots, p$) menjumlahkan bobot sinyal input persamaan (7)

$$Z_{in_{jk}} = v_{0j} + \sum_{i=1}^n x_i v_{ij} \quad (7)$$

dimana Z adalah neuron hidden; v_{0j} adalah neuron input ke- j ; x_i adalah neuron input ke- i ; v_{ij} adalah bobot neuron input ke neuron hidden. Dengan menerapkan fungsi aktivasi yang dihitung dengan persamaan (8)

$$Z_j = f(Z_{in_j}) \quad (8)$$

Dimana Z_j adalah unit ke- j pada lapisan tersembunyi; Z_{in_j} adalah keluaran untuk unit Z_j .

5. Tiap unit keluaran ($y_k, k = 1, \dots, m$) menjumlahkan input berbobot dengan menggunakan persamaan (9)

$$Y_{in_k} = w_{0k} + \sum_{j=1}^p Z_j w_{jk} \quad (9)$$

Dimana, Y_{in_k} adalah keluaran untuk unit Y_k ; w_{0k} adalah bobot bias untuk neuron hidden ke- k ; Z_j adalah unit ke- j pada lapisan tersembunyi; w_{jk} adalah bobot neuron hidden ke neuron output. Dengan menerapkan fungsi aktivasi yang dihitung dengan persamaan (10)

$$Y_k = f(Y_{in_k}) \quad (10)$$

6. Tiap Unit Keluaran ($y_k, k = 1, \dots, m$) menerima pola pelatihan masukan. Menghitung nilai *error* setiap layer dengan persamaan (11)

$$\delta_k = (t_k - y_k) f'(y_{in_k}) \quad (12)$$

Dimana, δ_k adalah faktor koreksi bobot w_{jk} ; t adalah target; y_k adalah neuron output ke- k ; y_{in_k} adalah keluaran untuk unit y_k . Perhitungan koreksi bobot dan bias-nya dilakukan dengan persamaan (13)

$$\begin{aligned} \Delta w_{jk} &= \alpha \delta_k x_j \\ \Delta w_{0k} &= \alpha \delta_k \end{aligned} \quad (13)$$

Dimana, Δw_{jk} adalah selisih antara $w_{jk}(t)$ dengan $w_{jk}(t+1)$; Δw_{0k} adalah bobot bias untuk neuron hidden ke- k ; α adalah learning rate; δ_k adalah faktor koreksi bobot w_{jk} ; x adalah input.

7. Tiap unit pada hidden layer ($z_j, j = 1, \dots, p$) menjumlahkan delta masukannya (dari unit-unit yang berada pada lapisan atasannya) dengan persamaan (14).

$$\delta_{in_j} = \sum_{k=1}^m \delta_k w_{jk} \quad (14)$$

Dimana, δ_k adalah faktor koreksi bobot w_{jk} ; w_{jk} adalah bobot neuron hidden ke neuron output. Perhitungan nilai *error* setiap layer dilakukan dengan persamaan (15)

$$\delta_j = \delta_{in_j} f'(x_{in_j}) \quad (15)$$

Dimana, δ_k adalah faktor koreksi bobot v_{ij} ; δ adalah faktor koreksi; x adalah input. Perhitungan koreksi bobot dan bias-nya dapat dilakukan dengan persamaan (16).

$$\Delta v_{ij} = \alpha \delta_j x_i \quad (16)$$

Dimana, Δv_{ij} adalah bobot neuron input ke neuron hidden; α adalah learning rate; δ_j adalah faktor koreksi bobot v_{ij} ; x_i adalah neuron input ke- i .

8. Tiap unit keluaran ($y_k, k = 1, \dots, m$) memperbaharui bobot dan bias ($j = 0, 1, \dots, p$) dihitung dengan persamaan (17).

$$w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta w_{jk} \quad (18)$$

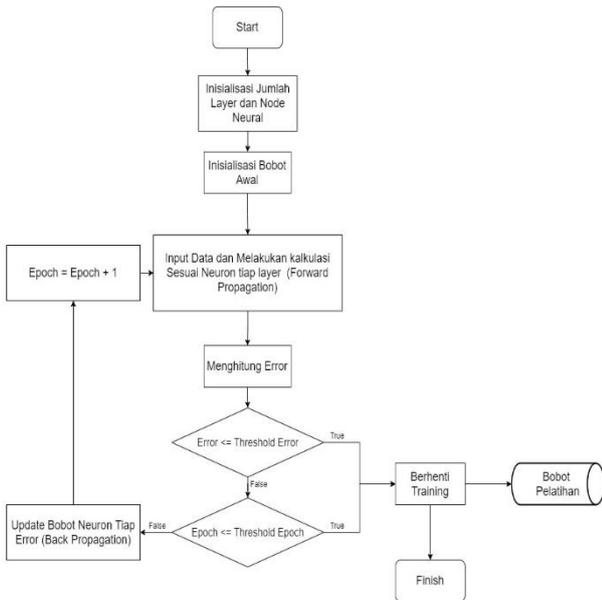
Dimana, w_{jk} adalah bobot neuron pada *hidden layer* ke *neuron output*; Δw_{jk} adalah selisih bobot neuron pada *hidden layer* ke *neuron output*. Tiap unit tersembunyi ($z_j, j = 1, \dots, p$) memperbaharui

bobot dan bias-nya ($i = 0,1, \dots n$) dihitung dengan persamaan (19)

$$v_{ij}(\text{baru}) = v_{ij}(\text{lama}) + \Delta v_{ij} \quad (19)$$

Dimana, v_{ij} adalah bobot neuron input ke neuron hidden; Δv_{ij} adalah selisih bobot neuron input ke neuron hidden.

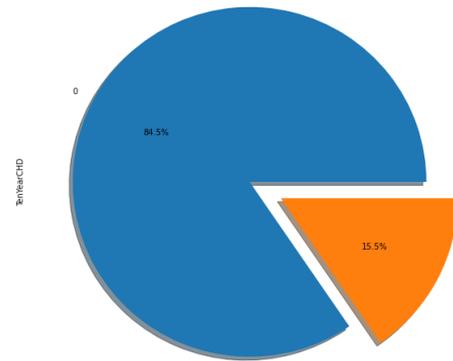
9. Syarat berhenti terpenuhi.



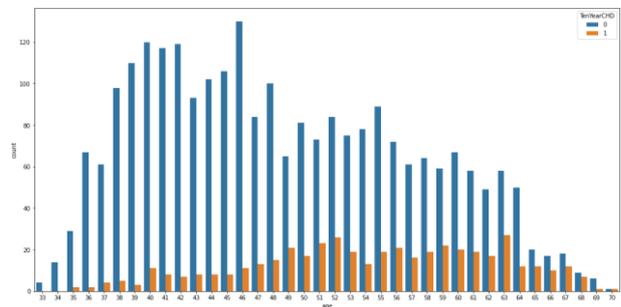
Gambar 3 Flowchart Back Propagation Neural Network

3. Hasil Percobaan

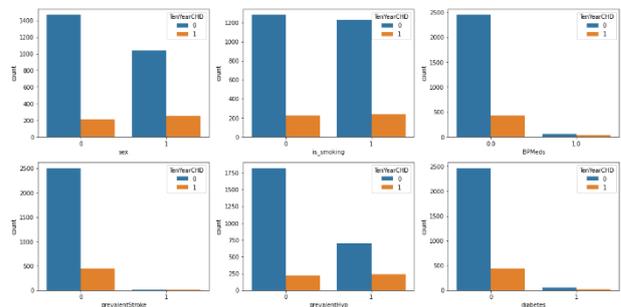
Pelatihan Model Back Propagation Neural Network pertama dilakukan dengan dataset yang disediakan dari situs Kaggle yang berjumlah 4238 data yang terdiri dari 2 kelas, yaitu berpotensi dan tidak berpotensi dengan pembagian 3594 kelas berpotensi dan 644 kelas tidak berpotensi. data ini kemudian di pecah menjadi menjadi 2968 data latih dan 1261 data uji. Data uji terdapat 185 data kelas berpotensi dan 1086 data kelas tidak berpotensi sedangkan data latih terdapat 459 data kelas berpotensi dan 2508 data kelas tidak berpotensi. Adapun visualisasi dataset sebagai berikut



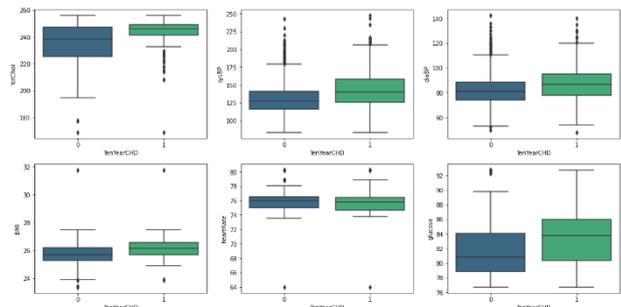
Gambar 4 Visualisasi Perbandingan Data Pengidap penyakit jantung koroner dan Bukan Pengidap penyakit jantung koroner



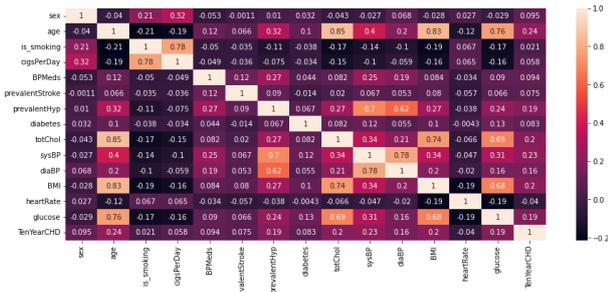
Gambar 5 Visualisasi Sebaran Pengidap penyakit jantung koroner dan Bukan Pengidap penyakit jantung koroner Berdasarkan Usia



Gambar 6 Visualisasi Sebaran Pengidap penyakit jantung koroner dan Bukan Pengidap penyakit jantung koroner pada Variable Jenis Kelamin, Kebiasaan Merokok, Aktif Konsumsi Obat Tekanan Darah, Memiliki Riwayat Penyakit Stroke, Memiliki Riwayat Penyakit Hipertensi dan Memiliki Riwayat Diabetes.



Gambar 7 Visualisasi Pengidap PJK dan Bukan Pengidap PJK Berdasarkan Data Total LDL Kolesterol, Tekanan Darah Sistolik, Tekanan Darah Diastolik, BMI, Rata-rata Detak Jantung dan Tingkat Glukosa Darah

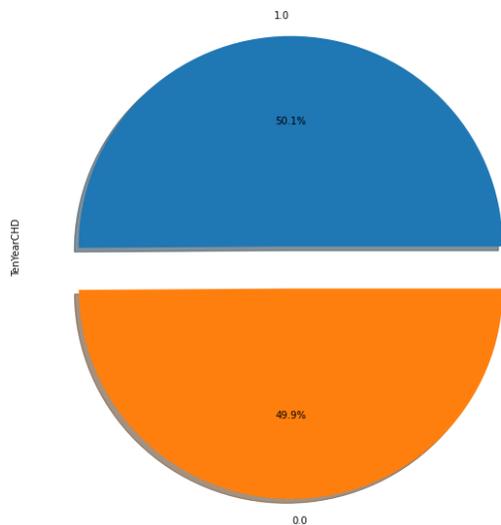


Gambar 8 Heatmap Hubungan Antar Variable

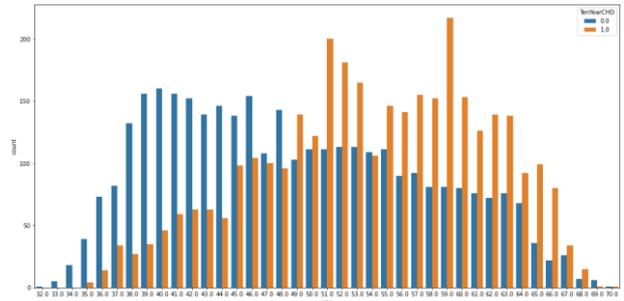
Berdasarkan dataset tersebut, pelatihan dilakukan dengan menerapkan *early stops*. *early stops* adalah strategi atau pendekatan untuk menghentikan training model untuk mencegah overfitting model. pendekatan yang dilakukan adalah dengan berhenti melakukan training saat nilai dari loss function telah mencapai batas dari *threshold* yang telah tentukan. pada pelatihan ini, *threshold* yang diberikan adalah jika *accuracy validation* tidak bertambah selama 32 epochs.

Dengan pendekatan pelatihan ini, model telah berhenti pada epoch ke 36, dengan mencapai accuracy hingga 0.8552, recall 0.0432 dan precision 0.5333. hasil pengujian pada data uji menunjukkan hasil yang tidak seimbang. hasil pengujian menunjukkan skor recall sangatlah kecil yaitu hanya pada 0.0432.

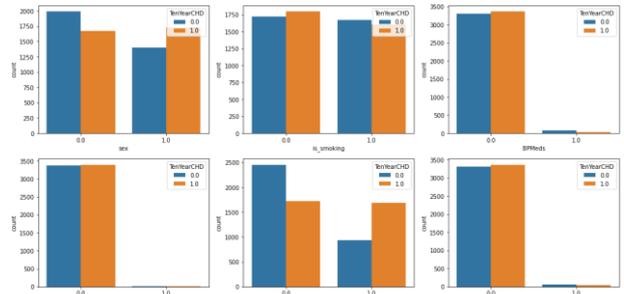
Pelatihan Model Back Propagation Neural Network kedua dilakukan pada dataset yang telah dilakukan SMOTE (*Synthetic Minority Oversampling Technique*). setelah menerapkan metode SMOTE ini, data uji yang disiapkan berjumlah 386 data dengan pembagian 193 data berpotensi mengidap penyakit jantung koroner dan 193 data tidak berpotensi mengidap penyakit jantung koroner, selain data uji, data latih yang dibuat berjumlah 6788 data dengan pembagian 3401 data berpotensi mengidap penyakit jantung koroner dan 3387 data tidak berpotensi mengidap penyakit jantung koroner. Sebaran datanya dapat dilihat pada visualisasi sebagai berikut.



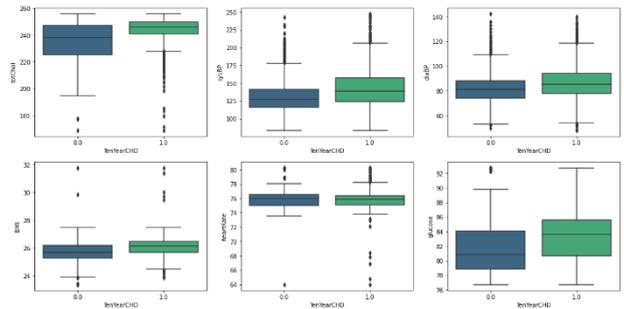
Gambar 9 Visualisasi Perbandingan Data Pengidap penyakit jantung koroner dan Bukan Pengidap penyakit jantung koroner



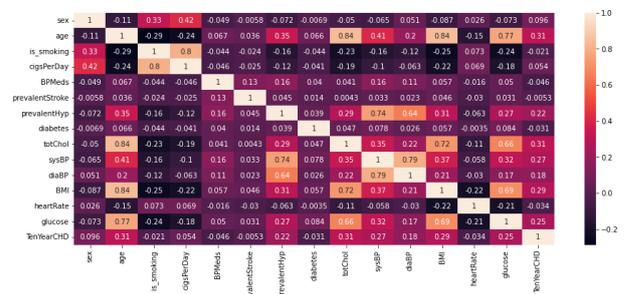
Gambar 10 Visualisasi Sebaran Pengidap Penyakit Jantung Koroner dan Bukan Pengidap Penyakit Jantung Koroner Berdasarkan Usia.



Gambar 11 Visualisasi Sebaran Pengidap penyakit jantung koroner dan Bukan Pengidap penyakit jantung koroner pada Variable Jenis Kelamin, Kebiasaan Merokok, Aktif Konsumsi Obat Tekanan Darah, Memiliki Riwayat Penyakit Stroke, Memiliki Riwayat Penyakit Hipertensi dan Memiliki Riwayat Diabetes



Gambar 12 Visualisasi Pengidap penyakit jantung koroner dan Bukan Pengidap penyakit jantung koroner Berdasarkan Data Total LDL Kolesterol, Tekanan Darah Sistolik, Tekanan Darah Diastolik, BMI, Rata-rata Detak Jantung dan Tingkat Glukosa Darah



Gambar 13 Heatmap Hubungan Antar Variable

Bentuk model yang digunakan dalam melatih dataset yang telah disampling sama dengan model yang digunakan untuk dataset dari situs Kaggle. hanya saja *early stops* pada pelatihan ini terjadi pada epoch ke 146 dengan mencapai *validation accuracy* hingga 0.5777, recall 0.4145 dan precision 0.6153. meskipun memiliki

validation accuracy dibawah model sebelumnya. model ini berhasil mencapai pengujian dengan nilai *recall* meningkat. dapat dilihat jika model dengan dataset yang melalui proses smote dapat menaikkan performa mendeteksi kejadian penyakit jantung koroner lebih baik.

Chairisni Lubis, memperoleh gelar Dra. dari Universitas Indonesia, Indonesia tahun 1989. Kemudian tahun 2000 memperoleh M. Kom., dari Universitas Indonesia, Indonesia. Saat ini sebagai Staf Pengajar program studi Teknik Informatika Universitas Tarumanagara.

Tabel 1 Perbandingan pelatihan sebelum dan sesudah SMOTE

Pelatihan	Epoch	Accuracy	Recall	Precision
1	36	0.8552	0.0432	0.5333
2	146	0.5777	0.4145	0.6153

4. Kesimpulan

Penerapan *Synthetic Minority Oversampling Technique (SMOTE)* dapat meningkatkan performa pelatihan model *back propagation neural network* dan meningkatkan performa model dalam memprediksi kelas minoritas.

Berdasarkan percobaan yang telah dilakukan, dengan penerapan beberapa hyperparameter, tidak didapatkan model yang memiliki performa baik pada data testing, hal ini juga dapat dikarenakan kurangnya ragam data pada kelas pengidap penyakit jantung koroner yang mengharuskan untuk melakukan sampling. Selain itu dapat dilihat pada sebaran dataset, jika terdapat beberapa feature yang memiliki range terlalu kecil. Seperti pada feature BMI hanya berkisar direntan 23.5 hingga 25.8 yang hanya termasuk di kategori healthy dan overweight tanpa adanya kategori underweight, obese dan extremely obese. Selain itu juga dapat dilihat pada feature Riwayat penyakit stroke dan Riwayat penyakit diabetes, terdapat perbedaan jumlah data yang sangat jauh antara data yang memiliki Riwayat maupun yang tidak memiliki Riwayat.

Pengembangan lebih lanjut dapat dilakukan dengan menambahkan jumlah data dengan kelas minoritas secara organik dan dapat menambahkan keragaman data agar memiliki range lebih luas.

REFERENSI

- [1] Who.int ‘Cardiovascular Diseases (CVDs)’, 2021. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- [2] Juslim, R Rukma dan Herawati, Fauna. “Penyakit Kardiovaskular”. Yogyakarta. Graha Ilmu, 2018.
- [3] Juslim, R Rukma dan Herawati, Fauna. “Penyakit Kardiovaskular”. Yogyakarta. Graha Ilmu, 2018.
- [4] Siringoringo, Rimbun. “Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote dan K-Nearest Neighbor”. Jurnal ISD. Vol. 3, Nomor 1, 2018
- [5] Russell, Stuart dan Norvig, Peter. “Artificial Intelligence A Modern Approach Third Edition”.New Jersey, Pearson Education. 2010.

Henri, menempuh Pendidikan S1 Teknik Informatika dengan gelar S.Kom dari Universitas Tarumanagara, Indonesia 2018.