

ANALISIS ALGORITMA KLASIFIKASI NEURAL NETWORK UNTUK DIAGNOSIS PENYAKIT KANKER PAYUDARA

Fitra Septia Nugraha¹; Muhammad Ja'far Shidiq²; Sri Rahayu³

Ilmu Komputer
STMIK Nusa Mandiri Jakarta
www.nusamandiri.ac.id

¹fitraseptia7@gmail.com; ²ash.shidiq.mj@gmail.com; ³srahayu110527@gmail.com

Abstract—One of the main causes of death is cancer. The most common cancer in women is breast cancer. This disease if it be known early could be overcome and even prevented. Data classification techniques could be used to predict which patients had breast cancer and not with some parameters. Using the Neural Network method and Rapid Miner 9.0 tools aims to predict breast cancer diagnosis and then produced an accuracy value of 71,83%, precision 81,08% and recall of 69,17% with AUC of 0,806 which means that the classification was good so that patients with parameters there could be predicted which ones were breast cancer patients and which were not, so this pattern could be used as a benchmark for diagnosis so that it could be detected earlier and was expected to reduce the number of deaths from breast cancer.

Keywords : breast cancer, classification, data mining, neural network, rapid miner.

Intisari—Salahsatu penyebab kematian utama adalah kanker. Kanker yang paling umum pada wanita adalah kanker payudara. Penyakit ini jika dapat diketahui sejak dini dapat ditanggulangi bahkan dicegah. Teknik datamining klasifikasi dapat digunakan untuk memprediksi pasien mana yang terkena kanker payudara dan tidak dengan beberapa parameter yang ada. Dengan menggunakan metode Neural Network dan tools Rapid Miner 9.0 bertujuan untuk memprediksi diagnosis kanker payudara dan kemudian menghasilkan nilai accuracy 71,83%, precision 81,08% dan recall 69,17% dengan AUC sebesar 0,806 yang artinya klasifikasi dinyatakan baik sehingga pasien dengan parameter yang ada dapat diprediksi mana saja yang merupakan pasien kanker payudara dan mana yang bukan, sehingga pola ini dapat digunakan sebagai tolak ukur diagnosis sehingga dapat dideteksi lebih dini dan diharapkan dapat menekan angka kematian akibat kanker payudara.

Kata Kunci: kanker payudara, klasifikasi, data mining, neural network, rapid miner.

PENDAHULUAN

Penyakit kanker merupakan salah satu penyebab kematian utama di seluruh dunia. Pada tahun 2012, sekitar 8,2 juta kematian disebabkan oleh kanker. Kanker paru, hati, perut, kolorektal, dan kanker payudara adalah penyebab terbesar kematian akibat kanker setiap tahunnya. (Pusdatin, 2015)

Kanker payudara adalah kanker yang paling umum pada wanita dan penyebab utama kematian kanker. WHO mengestimasikan bahwa 84 juta orang meninggal akibat kanker dalam rentang waktu 2005-2015. Survei yang dilakukan WHO tersebut menyatakan bahwa 8 sampai 9 persen wanita mengalami kanker payudara. Hal itu membuat kanker payudara sebagai jenis kanker yang paling banyak ditemui pada wanita setelah kanker leher rahim. Anggorowati dalam (Ma'arif & Arifin, 2017).

Hal tersebut didukung data statistik *American Cancer Society* yang menunjukkan bahwa kanker payudara menempati urutan pertama dengan jumlah kasus terbanyak dari seluruh jenis kasus tumor ganas (kanker) di seluruh dunia. *American Cancer Society* dalam (Handayani, Jamal, & Septiandri, 2018). Frekuensi kasus penyakit ini relatif tinggi di negara maju dan merupakan jenis kanker yang banyak diderita dari jenis kanker lainnya. Namun di Afrika dan Asia Tenggara, termasuk Indonesia, kanker payudara menempati peringkat kedua setelah kanker serviks. (Wahyuni, 2017). Menurut data terakhir WHO, angka kematian karena kanker payudara di Indonesia mencapai 1,41% dari seluruh kematian atau angka kematian disesuaikan dengan usia adalah 2.025 per 100.000 penduduk. deherba.com dalam (Wahyuni, 2017).

Penyakit kanker payudara atau breast cancer adalah penyakit yang ditandai dengan kelainan siklus sel khas yang menimbulkan kemampuan sel ini dapat menyebabkan pembelahan sel melebihi batas normal. Akibatnya pembelahan sel yang terjadi menyerang jaringan biologis di dekatnya, dan dapat bermigrasi ke

jaringan tubuh melalui sirkulasi darah. Sairun dalam (Ma'arif & Arifin, 2017). Meskipun etiologi kanker payudara tidak diketahui, faktor risiko berbagai kemungkinan mempengaruhi perkembangan penyakit ini termasuk faktor genetik, faktor karsinogen dan faktor perilaku atau gaya hidup.

Hingga saat ini, salah satu cara pengobatan kanker payudara yang umum dilakukan adalah dengan pembedahan dan jika perlu dilanjutkan dengan kemoterapi maupun radiasi. Akan tetapi pengobatan tersebut tidak akan memberikan dampak yang signifikan jika kankernya sudah mencapai stadium akhir. Hal ini terjadi karena dinegara berkembang seperti Indonesia, kanker payudara tidak didiagnosis sejak dini. Biasanya, kanker didiagnosis setelah memasuki stadium lanjut dan memiliki tingkat *survivability* rendah. Oleh karena itu apabila penyakit ini dapat dideteksi lebih awal, dampak buruk yang mungkin ditimbulkan oleh kanker payudara dapat dicegah. (Zamani, Amaliah, & Munif, 2012).

Terdapat beberapa penelitian yang sudah dilakukan sebelumnya dengan menggunakan data *Wisconsin Breast Cancer* (WBC) yang tersedia di *UCI Machine Learning Repository*, yang berisi 699 *clinical instances* mengenai data tumor ganas payudara dan tumor jinak payudara dari hasil diagnosis *breast masses* berdasarkan tes FNA. Namun penelitian tersebut hanya sebatas mengetahui normal atau tidaknya payudara, ganas atau tidaknya tumor di payudara dan hanya menguji variabel-variabel yang berpengaruh terhadap hasil klasifikasi tumor payudara, seperti riwayat keluarga, terapi hormon, ukuran tumor, letak tumor, dan usia menopause. Makhfudhoh dalam (Ma'arif & Arifin, 2017).

Pada umumnya pendeteksian tingkat keganasan kanker payudara adalah dengan cara prognosis. Prognosis adalah "tebakan terbaik" tim medis dalam menentukan sembuh atau tidaknya pasien dari kanker payudara. (Rachman & Purnami, 2012). Selain dengan prognosis, cara lainnya adalah pemanfaatan *bionformatic* dengan menggunakan teknik *data mining* (Salama, Abdelhalim, & Zeid, 2012), karena telah terbukti dapat mendeteksi tingkat keganasan kanker payudara.

Seiring dengan kemajuan teknologi informasi terutama dalam bidang kecerdasan buatan, teknik *machine learning* diperkenalkan untuk membantu meningkatkan kemampuan pendeteksian otomatis. Dengan bantuan sistem ini, kemungkinan akan kesalahan diagnosis yang dilakukan oleh para ahli medis dapat dihindari, dan data medis dapat diperiksa dalam kurun waktu yang singkat serta lebih rinci M. F. Akay dalam (Wahyuni, 2017).

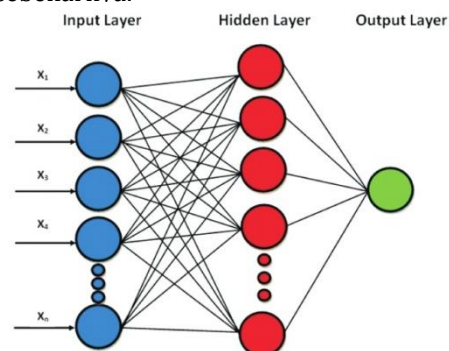
Data mining dapat diaplikasikan di bidang kesehatan misalnya mendiagnosis penyakit kanker payudara, penyakit jantung, penyakit diabetes dan lain-lain Larose dalam (Hermawanti, 2012). Salah satu teknik ini dalam hal klasifikasi adalah Neural Network yang menjadi sistem diagnosis otomatis untuk mendeteksi kanker payudara dengan tingkat klasifikasi yang benar dari sistem yang disusulkan yaitu 95,6%. (Karabatak & Ince, 2009). Artificial Neural Network juga pernah dibandingkan dengan Regresi Logistic dalam memprediksi kanker payudara dengan nilai AUCs lebih tinggi yaitu 0,965 dan Regresi Logistic 0,963. (Ayer et al., 2010).

Dari uraian di atas maka tujuan penelitian ini adalah untuk memprediksi atau mendiagnosis dini penyakit kanker payudara dengan menggunakan algoritma klasifikasi *machine learning neural network* sebagai proses *training*, *learning* dan *testingnya*. Parameter yang digunakan dalam penelitian ini berbeda dengan penelitian-penelitian sebelumnya yang mana dalam penelitian ini parameter yang digunakan adalah data Breast Cancer Coimbra yang berisi 116 *clinical instances* mengenai data pasien yang terkena kanker payudara berdasarkan tes darah, sehingga kanker dapat dideteksi sejak dini, sehingga diharapkan dapat segera ditanggulangi dan menekan angka kematian bukan dari bentuk kankernya itu sendiri seperti yang dilakukan peneliti sebelumnya. Untuk menentukan tingkat keberhasilan diagnosis kanker payudara dilakukan dengan 2 pengujian yaitu dengan menggunakan metode *confusion matrix* dan uji kehandalan ROC Curve.

BAHAN DAN METODE

Neural Network

Pada tahun 1943, McCulloch dan Pitts memperkenalkan model matematika yang merupakan penyederhanaan dari struktur sel saraf yang sebenarnya.



Sumber : (Khademi & Jamal, 2016)

Gambar 1. McCulloch & Pitts neuron model

Gambar diatas memperlihatkan bahwa sebuah neuron memiliki tiga komponen :

1. Synapse (w_1, w_2, \dots, w_n)^T
2. Alat penambah (adder)
3. Fungsi aktivasi (f)

Korelasi antara ketiga komponen ini dirumuskan pada persamaan (1).

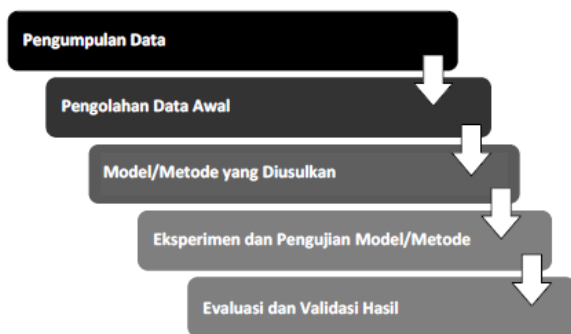
$$y = f(\sum_{i=1}^n x_i \times w_i) \dots \dots \dots (1)$$

Sumber : (Yahya, 2012)

Signal x berupa vektor berdimensi n (x_1, x_2, \dots, x_n)^T akan mengalami penguatan oleh synapse w (w_1, w_2, \dots, w_n)^T. Selanjutnya akumulasi dari penguatan tersebut akan mengalami transformasi oleh fungsi aktivasi f . Fungsi f ini akan memonitor, bila akumulasi penguatan signal itu telah melebihi batas tertentu, maka sel neuron yang semula berada dalam kondisi "0", akan mengeluarkan signal "1". Berdasarkan nilai output tersebut ($=y$), sebuah neuron dapat berada dalam dua status: "0" atau "1". Neuron disebut dalam kondisi *firing* bila menghasilkan output bernilai "1". (Yahya, 2012)

Metode Penelitian

Pada penelitian ini data yang digunakan adalah data penyakit kanker payudara (*Breast Cancer Coimbra*) yang didapat dari *Machine Learning Repository* UCI (Universitas California Invene) dengan alamat web: <http://archive.ics.uci.edu/ml/>. Dalam penelitian ini akan dilakukan beberapa langkah-langkah atau tahapan penelitian seperti gambar di bawah ini:



Sumber: (Septiani, 2017)
Gambar 2. Tahap Penelitian

1. Pengumpulan Data

Teknik pengumpulan data ialah teknik atau cara-cara yang dapat digunakan untuk menggunakan data Riduwan dalam (Septiani, 2017). Dalam pengumpulan data terdapat sumber data, sumber data yang dihimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan

apabila melalui tangan kedua disebut sumber sekunder Riduwan dalam (Septiani, 2017). Data pada penelitian ini merupakan data sekunder yang diperoleh dari Machine Learning Repository UCI (Universitas California, Invene) dengan alamat web <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra>. Data yang dikumpulkan adalah data antropometrik dan parameter yang dapat dikumpulkan dalam analisis darah rutin yang dapat digunakan sebagai biomarker kanker payudara. Data terkumpul sebanyak 116 data dengan 52 pasien dinyatakan sehat dan 64 pasien dinyatakan terdiagnosa kanker payudara dengan attribute age, BMI, Glucose, Insulin, Homa, Leptin, Adiponectin, Resistin, MCP.1 dan class.

2. Pengolahan Data Awal (*Pre Processing*)

Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan adalah sebagai berikut Vacellis dalam (Septiani, 2017):

- a. *Data validation*, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).
- b. *Data integration and transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penelitian ini bernilai numerikal.
- c. *Data size reduction and dicritization*, untuk memperoleh *dataset* dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informatif.

Dari proses pengolahan awal data di atas diperoleh sebanyak 116 data dengan 52 data dengan kelas "SEHAT" dan 64 data dengan kelas "PASIEN".

3. Metode yang diusulkan

Metode analisa merupakan metode yang diusulkan untuk mengetahui seberapa besar tingkat akurasi dari algoritma neural network dengan menggunakan komputasi rapid miner 9.0, hasil yang diperoleh dari analisa pengujian ini akan dijadikan sebagai dasar penentuan diagnosis kanker payudara.

4. Eksperimen dan Pengujian Methode

Data sebesar 116 akan dibagi menjadi 2 yaitu data untuk training set sebanyak 80% dari jumlah data dan testing set sebanyak 20% dari jumlah data. Kemudian data training di aplikasikan ke dalam metode, setelah itu dilakukan pengujian menggunakan model yang digunakan, memasukan data training ke dalam model dan menguji menggunakan Rapid Miner.

5. Evaluasi dan Validasi Hasil

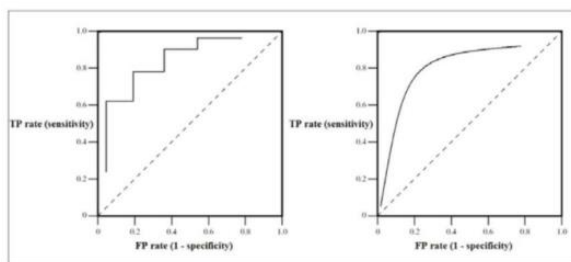
Evaluasi digunakan untuk melakukan pengamatan dan menganalisa hasil kerja neural network di rapid miner. Validasi dilakukan untuk melakukan pengukuran hasil prediksi.

a. Cross Validation

Merupakan pengujian standar yang dilakukan untuk memprediksi error rate. Setiap kelas pada data set harus diwakili dalam proporsi yang tepat antara data training dan testing. Data dibagi secara acak pada masing-masing kelas dengan perbandingan yang sama. Untuk mengurangi bias yang disebabkan oleh sampel tertentu, seluruh proses training dan pengujian diulangi beberapa kali dengan sampel yang berbeda. Tingkat kesalahan pada iterasi yang berbeda akan dihitung rata-ratanya untuk menghasilkan error rate secara keseluruhan (Hastuti, 2012).

b. Pengukuran menggunakan ROC Curve

ROC Curve adalah kurva ROC yang banyak digunakan para peneliti untuk menilai hasil prediksi. Kurva ROC menggambarkan kinerja klasifikasi tanpa memperhatikan distribusi kelas atau kesalahan, pada sumbu vertical menggambarkan nilai positif (TP) dan sumbu horizontal menandakan nilai negative (FP) (Sucipto, 2012). Lihat gambar berikut:



Sumber : (Sucipto, 2012)
Gambar 3. Contoh Grafik ROC

Beberapa hal penting yang perlu dicatat pada Gambar 2., pada titik kiri bawah (0,0) merupakan titik yang tidak pernah mengeluarkan klasifikasi positif, (1,1) terwakili titik kanan atas untuk menjelaskan nilai klasifikasi positif. Titik (0,1) merupakan nilai klasifikasi yang tidak terdapat nilai FN dan FP. Sedangkan garis diagonal yang membelah ruang ROC menggambarkan ruang diatas garis diagonal menandakan klasifikasi baik dan ruang dibawah garis diagonal menandakan klasifikasi buruk, sementara tebakan yang benar-benar acak terdapat pada sepanjang garis diagonal mulai dari kiri bawah sampai dengan kanan atas

(Sucipto, 2012). Sebuah metode umum untuk menghitung daeah dibawah kurva ROC adalah Area Under Curve (AUC) dimana bidang yang berada dibawah kurve mempunyai nilai yang selalu berada pada nilai 0,0 dan 1,0. Semakin tinggi luasnya maka akan semakin baik nilai klasifikasinya. Seperti petunjuk yang disajikan dalam (Sucipto, 2012) berikut ini:

- 0,90 – 1,00 = klasifikasi sangat baik
- 0,80 – 0,90 = klasifikasi baik
- 0,70 – 0,80 = klasifikasi rata-rata
- 0,60 – 0,70 = klasifikasi rendah
- 0,50 – 0,60 = kegagalan

c. Pengukuran menggunakan Confusion Matrix
Confusion matrix adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan (Indriani, 2014). Contoh confusion matrix untuk klasifikasi biner ditunjukkan pada tabel 1.

Tabel 1.
Confusion Matrix untuk klasifikasi biner

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Sumber: (Indriani, 2014)

Keterangan untuk Tabel 1. dinyatakan sebagai berikut:

- True Positive (TP), yaitu jumlah dokumen dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.
- True Negative (TN), yaitu jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.
- False Positive (FP), yaitu jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.
- False Negative (FN), yaitu jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

Perhitungan akurasi dinyatakan dalam persamaan 2.

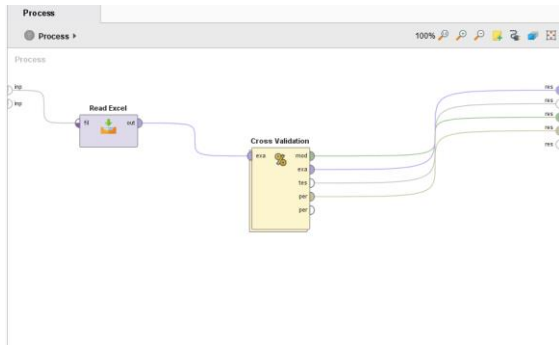
$$Accuracy = \frac{(TN+TP)}{(TN+FN+TP+FN)} \times 100 \%.....(2)$$

Sumber: (Indriani, 2014)

HASIL DAN PEMBAHASAN

Analisa Pengujian Algoritma Neural Network

Dalam tahap ini kita akan membuat model dengan menggunakan algoritma Neural Network, sama halnya dengan algoritma yang lain proses urutan-urutan pengujian algoritma Neural Network juga menggunakan cara yang sama dengan pengujian model lainnya dalam pengujian dengan rapid miner, model seperti pada gambar 4 dibawah ini:



Sumber : (Nugraha, Shidiq, & Rahayu, 2019)
Gambar 4. Desain Model Neural Network

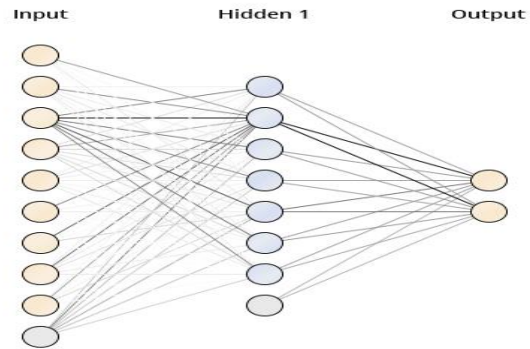
Dari desain model gambar 4., setelah data di import dan dimasukan dalam proses maka akan menghasilkan bentuk data seperti pada tabel 2. Data inilah yang akan digunakan untuk menghasilkan model dari metode Neural Network.

Tabel 2
Data Hasil Pengujian Neural Network

Age	BMI	Glucose	Insulin	Homa	Leptin	Adiponectin	Resistin	MCP.1	Label
48	23.50	70.00	2.71	0.47	8.81	9.70	8.00	417.11	0
83	20.69	92.00	3.12	0.71	8.84	5.43	4.06	468.79	0
82	23.13	91.00	4.50	1.01	17.94	22.43	9.28	554.70	0
68	21.37	77.00	3.23	0.61	9.88	7.17	12.77	928.22	0
86	21.11	92.00	3.55	0.81	6.70	4.82	10.58	773.92	0
49	22.85	92.00	3.23	0.73	6.83	13.68	10.32	530.41	0
89	22.70	77.00	4.69	0.89	6.96	5.59	12.94	1256.08	0
45	21.30	102.00	13.85	3.49	7.65	21.06	23.03	552.44	1
45	20.83	74.00	4.56	0.83	7.75	8.24	28.03	382.96	1
49	20.96	94.00	12.31	2.85	11.24	8.41	23.12	573.63	1
34	24.24	92.00	21.70	4.92	16.74	21.82	12.07	481.95	1
42	21.36	93.00	3.00	0.69	19.08	8.46	17.38	321.92	1
68	21.08	102.00	6.20	1.56	9.70	8.58	13.74	448.80	1
51	19.13	93.00	4.36	1.00	11.08	5.81	5.57	90.60	1
62	22.66	92.00	3.48	0.79	9.87	11.24	10.70	703.97	1
38	22.50	95.00	5.26	1.23	8.44	4.77	15.74	199.06	1
69	21.51	112.00	6.68	1.85	32.58	4.14	15.70	713.24	1
49	21.37	78.00	2.64	0.51	6.33	3.89	22.94	737.67	1
51	22.89	103.00	2.74	0.70	8.02	9.35	11.56	359.23	1
59	22.83	98.00	6.86	1.66	14.90	4.23	8.21	355.31	1
86	27.18	138.00	19.91	6.78	90.28	14.11	4.35	90.09	1

Sumber: (Patrício, 2018)

Dalam Tabel 2. Dapat memberikan informasi bahwa dari data yang dimasukan kedalam metode Neural Network dengan apply model pada rapid miner menunjukkan hasil prediksi yang akan menentukan apakah metode Neural Network yang digunakan dapat memberikan prediksi terbaik.



Sumber : (Nugraha et al., 2019)
Gambar 5. Model Hasil Pengujian Neural Network

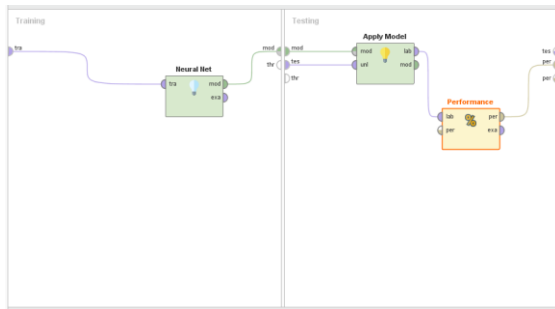
Dalam Gambar 5. Merupakan model hasil dari pengujian metode Neural Network yang dapat memberikan informasi bahwa dalam metode ini yang menjadi informasi penting adalah tentang hidden layer. Informasi dari model tersebut adalah pada sisi input terdapat 10 atribut yaitu age, BMI, glucose, insulin, homa, leptin, adiponectin, resistin, MCP.1 dan diagnosis, dimana atribut tersebut akan membentuk lapisan input (x) yang dapat diberi bobot $x_1, x_2, x_3, \dots, x_{10}$. Sedangkan pada posisi hidden layer juga akan membentuk lapisan tersembunyi (z), jarak antara lapisan input dengan lapisan tersembunyi mempunyai bobot yang disebut dengan matrik bobot pertama yang diberi inisial $v_{11}, v_{12}, \dots, v_{16}$. Selanjutnya pada posisi output akan muncul apa yang disebut lapisan output (y), dimana jarak antara (y) dan (z) dapat dihitung dengan menggunakan matrik bobot kedua yaitu $w_1, w_2, w_3, \dots, w_{10}$, setelah selesai perhitungan w maka akan dapat dihitung nilai outputnya dalam hal ini adalah sehat atau pasien.

Evaluasi dan Validasi

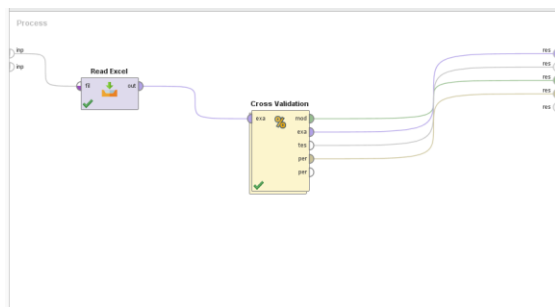
Penelitian ini bertujuan untuk menguji akurasi dari analisa diagnosis penyakit kanker payudara menggunakan algoritma neural network data yang dianalisa adalah data breast cancer coimbra yang di peroleh dari UCI Repository. Dengan bantuan rapid miner ternyata dapat diketahui algoritma neural network mempunyai akurasi yang tinggi.

Pengujian Model Neural Network

Pengujian model Neural Network dengan menggunakan testing dan validasi seperti ditunjukkan pada Gambar 6. dan Gambar 7. dapat menghasilkan nilai accuracy = 71,83%, precision =81,08% dan recall =69,17% dengan AUC sebesar =0,806 Seperti ditunjukkan pada gambar berikut ini:



Sumber : (Nugraha et al., 2019)
Gambar 6. Model Testing *Neural Network*



Sumber : (Nugraha et al., 2019)
Gambar 7. Desain Validasi *Neural Network*

accuracy: 71.83% +/- 19.76% (micro average: 72.41%)

	true 0	true 1	class precision
pred 0	40	20	66.67%
pred 1	12	44	78.57%
class recall	76.92%	68.75%	

Sumber : (Nugraha et al., 2019)
Gambar 8. Nilai *Accuracy* dari *Neural Network*

Berdasarkan Gambar 8. Dapat kita ketahui bahwa hasil *accuracy* dari metode klasifikasi *Neural Network* sebesar = 71,83% ini menunjukkan bahwa hasil akurasi yang diperoleh termasuk kedalam kategori baik.

precision: 81.08% +/- 23.55% (micro average: 78.57%) (positive class: 1)

	true 0	true 1	class precision
pred 0	40	20	66.67%
pred 1	12	44	78.57%
class recall	76.92%	68.75%	

Sumber : (Nugraha et al., 2019)
Gambar 9. Nilai *Precision* dari *Neural Network*

Berdasarkan Gambar 9. Diperoleh bahwa hasil *precision* dari metode klasifikasi *Neural Network* sebesar = 81,08 ini menunjukkan bahwa hasil akurasi yang diperoleh masuk dalam kategori baik.

recall: 69.17% +/- 27.02% (micro average: 68.75%) (positive class: 1)

	true 0	true 1	class precision
pred 0	40	20	66.67%
pred 1	12	44	78.57%
class recall	76.92%	68.75%	

Sumber : (Nugraha et al., 2019)
Gambar 10. Nilai *Recall* dari *Neural Network*

Berdasarkan Gambar 10. Diperoleh bahwa hasil *recall* dari metode klasifikasi *Neural Network* sebesar = 69,17% ini menunjukkan bahwa hasil akurasi yang diperoleh masuk dalam kategori cukup baik.

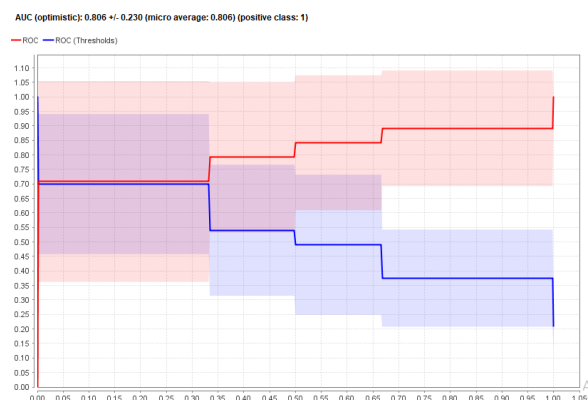
PerformanceVector

```

PerformanceVector:
accuracy: 71.83% +/- 19.76% (micro average: 72.41%)
ConfusionMatrix:
True: 0 1
0: 40 20
1: 12 44
AUC (optimistic): 0.806 +/- 0.230 (micro average: 0.806) (positive class: 1)
AUC (pessimistic): 0.806 +/- 0.230 (micro average: 0.806) (positive class: 1)
precision: 81.08% +/- 23.55% (micro average: 78.57%) (positive class: 1)
ConfusionMatrix:
True: 0 1
0: 40 20
1: 12 44
recall: 69.17% +/- 27.02% (micro average: 68.75%) (positive class: 1)
ConfusionMatrix:
True: 0 1
0: 40 20
1: 12 44
    
```

Sumber : (Nugraha et al., 2019)
Gambar 11. *Performance Vector Neural Network*

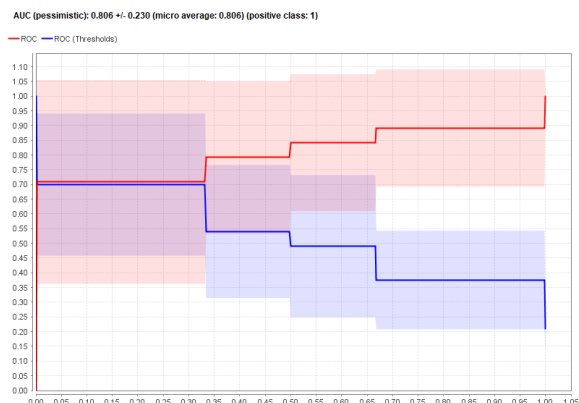
Gambar 11. Memberikan informasi secara keseluruhan tentang hasil dari validasi dengan metode klasifikasi *Neural Network*, disamping hasil *accuracy*, *precision* dan *recall*, juga terdapat nilai AUC dan informasi penting dalam performance vector ini yaitu *confusion matrix* pada masing-masing kategori sehat dan pasien dengan nilai jelas.



Sumber : (Nugraha et al., 2019)
Gambar 12. Hasil *AUC Optimistic* dari *Neural Network*

Gambar 12. menggambarkan grafik area under curve (AUC) *optimistic* hasil validasi *Neural*

Network dengan nilai yang tertera dalam gambar sebesar= 0,806.

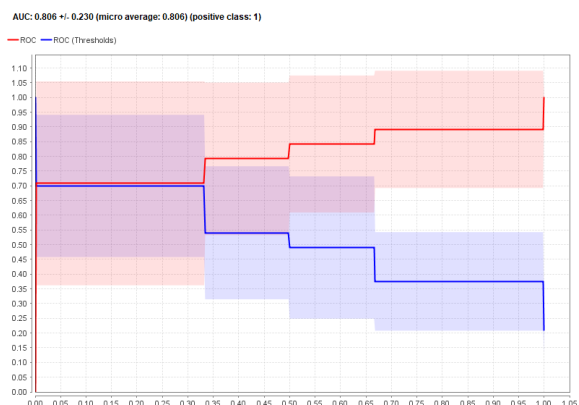


Sumber : (Nugraha et al., 2019)

Gambar 13.

AUC Pesimistic dari Neural Network

Gambar 13. Menggambarkan grafik area under curve (AUC) pessimistic hasil validasi Neural Network dengan hasil nilai dalam gambar sebesar= 0,806.



Sumber : (Nugraha et al., 2019)

Gambar 14. AUC Neural Network

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan menggunakan tools rapid miner 9.0 terhadap dataset breast cancer coimbra yang didapat dari UCI Repository yang diuji dengan metode Neural Network menghasilkan nilai accuracy = 71,83%, precision =81,08% dan recall =69,17% dengan AUC sebesar =0,806 yang menunjukkan bahwa hasil klasifikasinya baik, sehingga pasien dengan parameter yang ada dapat diprediksi mana saja yang merupakan pasien kanker payudara dan mana yang bukan, sehingga pola ini dapat digunakan sebagai tolak ukur diagnosis sehingga dapat dideteksi lebih dini dan

diharapkan dapat menekan angka kematian akibat kanker payudara.

REFERENSI

Ayer, T., Jagpreet Chhatwal, P., Oguzhan Alagoz, P., Charles E. Kahn, Jr, MD, M., Ryan W. Woods, MD, M., S., E., & Burnside, MD, MPH, M. (2010). *Informatics in Radiology Comparison of Logistic Regression and Artificial Neural Network Models in.*

Handayani, A., Jamal, A., & Septiandri, A. A. (2018). Evaluasi Tiga Jenis Algoritme Berbasis Pembelajaran Mesin untuk Klasifikasi Jenis Tumor Payudara. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*. <https://doi.org/10.22146/jnteti.v6i4.350>

Hastuti, K. (2012). *Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif. 2012*(Semantik), 241–249.

Hermawanti, L. (2012). Penerapan algoritma klasifikasi c4.5 untuk diagnosis penyakit kanker payudara. *Jurnal Teknik UNISFAT*, 7(1), 57–64.

Indriani, A. (2014). Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta, 21 Juni 2014, ISSN: 1907*, 5–10.

Karabatak, M., & Ince, M. C. Expert Systems with Applications An expert system for detection of breast cancer based on association rules and neural network. , 36 *Expert Systems With Applications* § (2009).

Khademi, F., & Jamal, S. M. (2016). RESEARCH PAPERS PREDICTING THE 28 DAYS COMPRESSIVE STRENGTH OF. *I-Manager's Journal on Civil Engineering*, 6(August), 1–7. <https://doi.org/10.26634/jce.6.2.5936>

Ma'arif, F., & Arifin, T. (2017). Optimasi Fitur Menggunakan Backward Elimination Dan Algoritma SVM Untuk Klasifikasi Kanker Payudara. *JURNAL INFORMATIKA*.

Nugraha, F. S., Shidiq, M. J., & Rahayu, S. (2019). ANALISIS ALGORITMA KLASIFIKASI NEURAL NETWORK UNTUK. *Jurnal Pilar Nusa Mandiri*, 15(1), 1–7.

Patrício, M. (2018). *Breast Cancer Coimbra Data Set* (Vol. 1).

Pusdatin. (2015). *Situasi penyakit kanker 4*. Kementerian Kesehatan RI. Jakarta.

Rachman, F., & Purnami, W. (2012). Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi Logistik Ordinal Dan Support Vector Machine (SVM). *Jurnal Sains Dan Seni Its*.

Salama, G. I., Abdelhalim, M. B., & Zeid, M. A. E. (2012). Experimental comparison of classifiers for breast cancer diagnosis. *Proceedings - ICCES 2012: 2012 International Conference on Computer Engineering and Systems*.
<https://doi.org/10.1109/ICCES.2012.6408508>

Sucipto, A. (2012). CREDIT PREDICTION WITH NEURAL NETWORK ALGORITHM. *PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI_U)*, 978-979-36(15), 1-10.

Wahyuni, E. S. (2017). PENERAPAN METODE SELEKSI FITUR UNTUK MENINGKATKAN HASIL DIAGNOSIS KANKER PAYUDARA. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*.
<https://doi.org/10.24176/simet.v7i1.516>

Yahya, S. (2012). *Fuzzy Logic , Neural Network , Genetic Algorithm Knowledge Based Expert System and Computational Intelligence*.

Zamani, A. M., Amaliah, B., & Munif, A. (2012). Implementasi Algoritma Genetika pada Struktur Backpropagation Neural Network untuk Klasifikasi Kanker Payudara. *JURNAL TEKNIK ITS, Vol. 1 ISS*.