

Penerapan Algoritma Neural Network untuk Klasifikasi Kanker Paru

Evy Priyanti

Universitas Bina Sarana Informatika

evy.evp@bsi.ac.id

Abstrak - Algoritma Neural Network merupakan algoritma yang digunakan mempelajari cara kerja otak manusia yang diterapkan dengan neuron yang dihubungkan dengan milyaran jaringan syaraf dan mampu bekerja dalam banyak proses pembelajaran data, dalam hal ini algoritma neural network akan mempelajari klasifikasi pada kanker paru-paru. Kanker paru merupakan jenis kanker terbesar ketiga di Indonesia. Kanker paru ini akan dibedakan menjadi tiga class kanker paru secara patologis. Pada class 1 terdiri dari 9 observasi, class 2 terdiri dari 13 observasi dan class 3 terdiri dari 10 observasi. Sulitnya menentukan jenis kanker berdasarkan patologisnya membuat neural network mampu menerapkan konsep neuron untuk mengklasifikasi patologi dari kanker paru tersebut. Hasil dari pengujian kanker paru dengan algoritma neural network ini menghasilkan nilai akurasi sebesar 75%, yaitu pada class kanker 1 sebesar 85% ini membuktikan bahwa kanker class 1 mendominasi hasil dari klasifikasi kanker paru, dengan demikian maka algoritma neural network dapat dipastikan mengklasifikasi patologi pada kanker paru yang didapat dari penelitian survey oleh Stefan Aeberhard dengan jumlah sample sebanyak 32 orang dan jumlah atribut sebanyak 57 buah, satu buah atribut class label dan 56 atribut berupa nominal integer dengan batasan antara 0-3 class pada UCI Dataset dengan class 1 sebagai kanker terbanyak yaitu sebesar 85% dengan nilai akurasi sebesar 75%.
;Kata Kunci : Neural Network, Kanker Paru, Akurasi

Abstract - *Neural Network Algorithm is an algorithm that is used to study the workings of the human brain which is applied to neurons connected to billions of network requirements and is able to work in many data learning processes, in this case the neural network algorithm will study the classification of lung cancer. Lung cancer is the third largest type of cancer in Indonesia. Lung cancer is divided into three classes of pathological lung cancer. Class 1 consists of 9 observations, class 2 consists of 13 observations and class 3 consists of 10 observations. The difficulty in determining the type of cancer based on its pathology makes the neural network able to apply the concept of neurons to classify the pathology of lung cancer. The results of lung cancer testing with this neural network algorithm produce an accuracy value of 75%, namely in cancer class 1 of 85% this proves that class 1 cancer dominates the results of lung cancer classification, thus the neural network algorithm can be ascertained in classifying the pathology of cancer. lungs obtained from survey research by Stefan Aeberhard with a sample size of 32 people and a total of 57 attributes, one class label attribute and 56 attributes in the form of a nominal integer with a limit between 0-3 classes on the UCI dataset with class 1 as the most cancer, namely equal to 85% with an accuracy value of 75%.*

Keywords: Neural Network, Lung Cancer, Accuracy

I. PENDAHULUAN

Neural Network merupakan cara dalam mengenali hubungan sekumpulan data melalui proses pembelajaran cara kerja otak manusia yaitu jaringan syaraf yang mengacu pada neuron baik yang organik maupun buatan. Dengan jaringan syaraf inilah proses pembelajaran dalam neural network bekerja. Pembelajaran yang dilakukan neural network sangat bergantung dengan banyaknya neuron yang masuk sebagai landasan dalam proses pembelajaran atau learning.

Kanker paru merupakan jaringan abnormal pada jaringan paru (Elizabeth, 2021).

Kanker paru biasanya ditemui oleh seseorang yang biasa merokok dan memiliki pola hidup tidak sehat, kanker paru juga merupakan jenis kanker terbesar ketiga di Indonesia. Walaupun terkadang kanker paru juga ditemui pada pasien yang tidak merokok akan tetapi sering

terpapar zat kimia berbahaya atau sering menghirup asap rokok dari orang lain atau peroko pasif. Untuk mengobati kanker paru ini jika masih dalam tahap awal tingkat keberhasilan akan tinggi, akan tetapi kanker paru ini di tahap awal tidak menunjukkan tanda-tanda atau gejala pada tahap awal. Beberapa gejala kanker paru yang biasa terjadi yaitu dengan adanya batuk yang sudah kronis bahkan dengan batuk berdarah, penurunan berat badan yang sangat drastis, rasa nyeri pada dada dan tulang yang disertai dengan sesak nafas.

Beberapa factor yang dapat memicu terjadinya kanker paru diantaranya adalah memiliki anggota keluarga dengan pasien kanker paru, tinggal atau bekerja di lingkungan dengan udara yang tidak sehat atau tercemar zat berbahaya atau polusi udara, dan pernah menjalani radioterapi (Willy, 2019).

II. METODOLOGI PENELITIAN

Penelitian kali ini menggunakan data sekunder dengan metode penelitian survey oleh Stefan Aeberhard dengan jumlah sample sebanyak 32 orang dan jumlah atribut sebanyak 57 buah, satu buah atribut class label dan 56 atribut berupa nominal integer dengan batasan antara 0-3 yang didapat dari Uci dataset (Aeberhard, 1991).

Pada penelitian ini digunakan neural network dalam klasifikasi jenis kanker paru. Pada dasarnya neural network bekerja dengan beberapa tahapan diantaranya (AG Farizawani, 2019):

- a. proses inisialisasi
- b. proses aktivasi
- c. Latihan bobot atau beban
- d. Iterasi atau pengulangan yang digunakan dalam tugas klasifikasi

Neural network merupakan algoritma yang dirancang untuk mengenali pola yang dapat mereka kenali dari vector dimana data diseluruh dunia baik berupa gambar, suara, teks atau deret waktu (Nicholson, n.d.)

Jaringan saraf bekerja mirip dengan jaringan saraf otak manusia. Sebuah "neuron" dalam jaringan saraf adalah fungsi matematika yang mengumpulkan dan mengklasifikasikan informasi menurut arsitektur tertentu. Jaringan ini memiliki kemiripan yang kuat dengan metode statistik seperti pemasangan kurva dan analisis regresi (Xiongwen Pang, 2018).

Operator ini mempelajari model melalui jaringan neural umpan maju yang dilatih oleh algoritme propagasi mundur (perceptron multi-layer). Paragraf yang akan datang menjelaskan ide-ide dasar tentang jaringan saraf, jaringan saraf perlu-maju, propagasi balik dan multi-layer perceptron.

Jaringan syaraf tiruan (JST), biasa disebut jaringan syaraf tiruan (NN), adalah model matematis atau model komputasi yang terinspirasi dari struktur dan aspek fungsional jaringan saraf tiruan. Jaringan saraf terdiri dari sekelompok neuron buatan yang saling berhubungan, dan memproses informasi menggunakan pendekatan koneksionis untuk komputasi (prinsip koneksionis sentral adalah bahwa fenomena mental dapat dijelaskan oleh jaringan yang saling berhubungan dari unit sederhana dan sering seragam). Dalam kebanyakan kasus, JST adalah sistem adaptif yang mengubah strukturnya berdasarkan informasi eksternal atau internal yang mengalir melalui jaringan selama fase pembelajaran. Jaringan saraf modern biasanya digunakan untuk memodelkan hubungan yang kompleks antara masukan dan keluaran atau untuk menemukan pola dalam data (RapidMiner,

2020).

Jaringan saraf umpan maju adalah jaringan saraf tiruan di mana koneksi antar unit tidak membentuk siklus terarah. Dalam jaringan ini, informasi bergerak hanya dalam satu arah, maju, dari node input, melalui node tersembunyi (jika ada) ke node output. Tidak ada siklus atau putaran dalam jaringan.

Algoritma propagasi mundur adalah metode pembelajaran terbimbing yang dapat dibagi menjadi dua tahap: propagasi dan pembaruan bobot. Kedua fase tersebut diulangi hingga kinerja jaringan cukup baik. Dalam algoritma propagasi balik, nilai keluaran dibandingkan dengan jawaban yang benar untuk menghitung nilai dari beberapa fungsi kesalahan yang telah ditentukan. Dengan berbagai teknik, kesalahan kemudian diumpungkan kembali melalui jaringan. Dengan menggunakan informasi ini, algoritme menyesuaikan bobot setiap koneksi untuk mengurangi nilai fungsi kesalahan dengan sejumlah kecil. Setelah mengulangi proses ini untuk sejumlah besar siklus pelatihan, jaringan biasanya akan menyatu ke beberapa keadaan di mana kesalahan penghitungannya kecil. Dalam kasus ini, dapat dikatakan bahwa jaringan telah mempelajari fungsi target tertentu (RapidMiner, 2020).

Multilayer perceptron (MLP) adalah model jaringan saraf tiruan umpan maju yang memetakan kumpulan data masukan ke dalam satu set keluaran yang sesuai. MLP terdiri dari beberapa lapisan node dalam grafik berarah, dengan setiap lapisan terhubung sepenuhnya ke lapisan berikutnya. Kecuali untuk node input, setiap node adalah neuron (atau elemen pemrosesan) dengan fungsi aktivasi nonlinier. MLP menggunakan propagasi balik untuk melatih jaringan. Kelas jaringan ini terdiri dari beberapa lapisan unit komputasi, biasanya saling berhubungan dengan cara umpan-maju. Dalam banyak aplikasi, unit jaringan ini menerapkan fungsi sigmoid sebagai fungsi aktivasi.

Pada operator ini fungsi sigmoid biasa digunakan sebagai fungsi aktivasi. Oleh karena itu, rentang nilai atribut harus diskalakan ke -1 dan +1. Ini dapat dilakukan melalui parameter normalisasi. Jenis simpul keluaran adalah sigmoid jika data pembelajaran mendeskripsikan tugas klasifikasi dan linier jika data pembelajaran mendeskripsikan tugas regresi numerik.

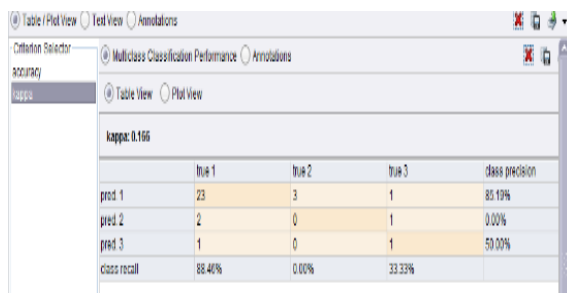
III. HASIL DAN PEMBAHASAN

Hasil dari penelitian tentang kanker paru menggunakan Neural Network memperoleh hasil akurasi sebesar 75% dengan 4 data pertama sebagai berikut:

Tabel 1. Dataset Kanker Paru

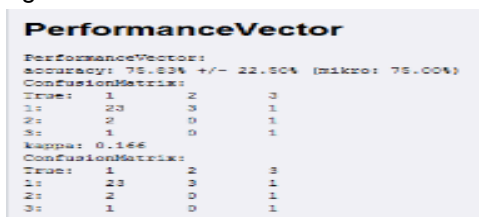
class	atribut 1	atribut ...	atribut 56
1	0	...	0
1	0	...	3
1	0	...	3
1	0	...	3
1	0	...	2
1	0	...	3
1	0	...	2
1	0	...	2
1	0	...	1

Pada tabel 1 adalah daftar tabel yang terdiri dari dataset kanker paru dari Stefan Aeberhard dengan jumlah sample sebanyak 32 orang dan jumlah atribut sebanyak 57 buah, satu buah atribut class label dan 56 atribut berupa nominal integer dengan batasan antara 0-3 yang didapat dari Uci dataset (Aeberhard, 1991).



Gambar 1 Plot View

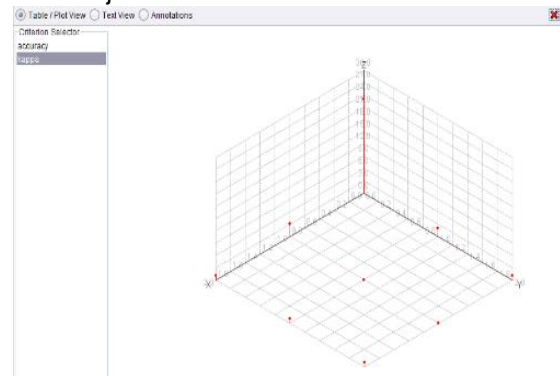
Pada Gambar 1 terlihat nilai kappa sebesar 0.166. nilai kappa adalah ukuran yang menyatakan konsistensi pengukuran yang dilakukan dua orang penilai (Rater) atau konsistensi antar dua metode pengukuran atau dapat juga mengukur konsistensi antar dua alat pengukuran



Gambar 2 Performance Vector

Terlihat bahwa hasil pengujian mendapatkan nilai akurasi sebesar 75.03% dengan confusionMatrix true untuk output 1 sebesar 23, output 2 sebesar 3 dan output 3 sebesar 1 ini membuktikan bahwa pada class 1 dimana hasil 85% berada pada class tersebut merupakan class yang terbesar atau terbanyak penderita kanker paru dimana algoritma neural network membuktikannya dengan menghasilkan nilai

akurasi sebesar 75% dan ini sangat membantu dalam menentukan pasien dalam penanganan lebih lanjut.



Gambar 3 Grafik Kappa

Pada Gambar 3.3 grafik Kappa ini terlihat koordinat dimana setiap node atau Sigmoid pada setiap hidden layer yang ada. Untuk dapat melihat tingkat validasi dari penelitian dengan pendekatan kappa dan realibilitas hasil penelitian ini. (Napitupulu, 2014).

Berikut hasil dari pengujian kanker paru-paru dengan neural network :

Tabel 2 ImprovedNeuralNet

Node 1 (Sigmoid)	Node...(Sigmoid)	Node 5 (Sigmoid)
atribut 1: 0.026	atribut 1 :	atribut 1: 0.041
atribut 2: 0.032	atribut 2 :	atribut 2: 0.142
atribut 3: -0.351	atribut 3 :	atribut 3: -0.357
atribut 4: 0.106	atribut 4 :	atribut 4: 0.149
atribut 5: -0.183	atribut 5 :	atribut 5: -0.010
...
atribut 56: 0.041	...	atribut 56: 0.031
Bias: 0.045	Bias :	Bias: 0.010

Pada tabel 2 ImprovedNeuralNet menentukan node 1 hingga node 5 yang terdiri dari hidden layer 1 dengan nilai bias pada sigmoid pada node 1 yaitu 0.045, node 2 dengan nilai bias 0,031 node 3 dengan nilai bias -0.036, node 4 dengan nilai bias 0,076 dan nilai bias dari node 5 sebesar 0,010.

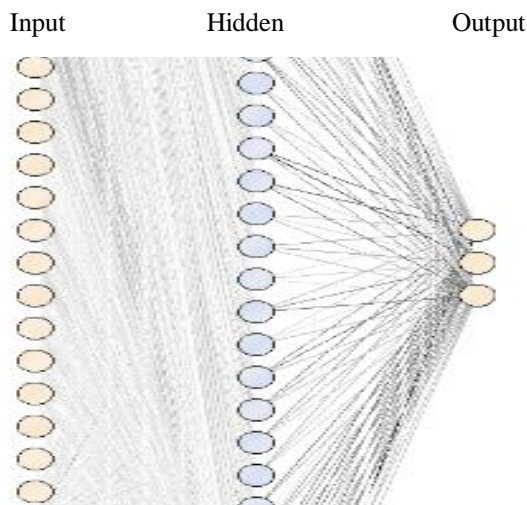
Tabel 3. Output

Class '1' (Sigmoid)	Class '2' (Sigmoid)	Class '3' (Sigmoid)
Node 1: 0.513	Node 1: -0.007	Node 1: -0.830
Node 2: -0.335	Node 2: 1.313	Node 2: -1.232
Node 3: 0.264	Node 3: -0.216	Node 3: -0.529
Node 4: -0.598	Node 4: 0.116	Node 4: 0.135
Node 5: 0.760	Node 5: -0.389	Node 5: -0.711
...
Node 31: -0.504	Node 31: 0.737	Node 31: -0.533
Threshold: -1.990	Threshold: 0.642	Threshold: 0.524

Pada tabel 3 terlihat output dari hasil yang diinginkan yaitu class 1 dengan node 1 hingga 31 threshold sebesar -1.990 dan class 2 terdapat threshold sebesar 0.642 dan class 3 terdapat 0.524 threshold.

Pada table ImprovedNeuralNet membuktikan nilai yang didapat dari hasil penelitian yang dilakukan dengan menggunakan algoritma neural network pada penderita kanker paru dimana pada class 1 merupakan penderita paling banyak yaitu threshold sebesar -1.990 dan merupakan jenis patologis yang paling banyak dijumpai pada hasil penelitian yang terdiri dari sample sebanyak 32 orang dan jumlah atribut sebanyak 57 buah, satu buah atribut class label dan 56 atribut berupa nominal integer dengan batasan antara 0-3 yang didapat dari Uci dataset.

Kebanyakan pasien adalah pasien lama yang sebenarnya mereka tidak sadar akan kanker yang mereka derita dikarenakan gejala dari kanker paru pada stadium awal yaitu batuk yang berkelanjutan hingga batuk berdarah, sesak nafas hingga nyeri dada, suara serak hingga suara mengik saat menarik atau menghembuskan nafas penurunan berat badan.



Gambar 4 Arsitektur Jaringan yang didapat dari Hasil Eksperimen Neural Network

Bagian dari arsitektur jaringan yang didapat dari hasil eksperimen neural network terdiri dari input, hidden dan output.

Input terdiri dari Training set

Port input mengharapkan ExampleSet. Ini adalah keluaran dari operator Retrieve dalam proses contoh kita. Output dari operator lain juga dapat digunakan sebagai input.

Output terdiri dari tiga class dimana setiap class terdiri dari 31 node.

Model (Improved Neural Net) yaitu Model Neural Net dikirim dari port keluaran ini. Model ini sekarang dapat diterapkan pada kumpulan

data yang tidak terlihat untuk prediksi atribut label.

Example set adalah ExampleSet yang diberikan sebagai masukan diteruskan tanpa mengubah ke keluaran melalui port ini. Ini biasanya digunakan untuk menggunakan kembali ExampleSet yang sama di operator lebih lanjut atau untuk melihat ExampleSet di Hasil Ruang Kerja.

Parameter

Hidden Layer adalah Parameter ini menjelaskan nama dan ukuran semua lapisan tersembunyi. Pengguna dapat menentukan struktur jaringan saraf dengan parameter ini. Setiap entri daftar menggambarkan lapisan tersembunyi baru. Setiap entri membutuhkan nama dan ukuran lapisan tersembunyi. Nama lapisan dapat dipilih secara sewenang-wenang. Ini hanya digunakan untuk menampilkan model. Perhatikan bahwa jumlah node sebenarnya akan menjadi satu lebih dari nilai yang ditentukan sebagai ukuran lapisan tersembunyi karena node konstan tambahan akan ditambahkan ke setiap lapisan. Node ini tidak akan terhubung ke lapisan sebelumnya. Jika nilai ukuran lapisan tersembunyi disetel ke -1, ukuran lapisan akan dihitung dari jumlah atribut kumpulan contoh masukan. Dalam hal ini, ukuran lapisan akan diatur ke $(\text{jumlah atribut} + \text{jumlah kelas}) / 2 + 1$. Jika pengguna tidak menentukan lapisan tersembunyi, lapisan tersembunyi default dengan jenis dan ukuran sigmoid sama dengan $(\text{jumlah atribut} + \text{jumlah kelas}) / 2 + 1$ akan dibuat dan ditambahkan ke internet. Jika hanya satu lapisan node yang ditentukan, node input terhubung langsung ke node output dan tidak ada lapisan tersembunyi yang akan digunakan.

training_cycles=Parameter ini menentukan jumlah siklus pelatihan yang digunakan untuk pelatihan jaringan saraf. Dalam propagasi balik, nilai keluaran dibandingkan dengan jawaban yang benar untuk menghitung nilai dari beberapa fungsi kesalahan yang telah ditentukan. Kesalahan tersebut kemudian diumpungkan kembali melalui jaringan. Dengan menggunakan informasi ini, algoritme menyesuaikan bobot setiap koneksi untuk mengurangi nilai fungsi kesalahan dengan sejumlah kecil. Proses ini diulangi sebanyak n kali. n dapat ditentukan menggunakan parameter ini. Range: integer

learning_rate=Parameter ini menentukan seberapa banyak kita mengubah bobot pada setiap langkah. Ini tidak boleh 0. Range: real
Momentum=Momentum tersebut hanya menambah sebagian kecil dari pembaruan bobot sebelumnya dengan yang sekarang. Ini mencegah maksima lokal dan memperhalus arah pengoptimalan. Range: real

Hasil dari penelitian yang dilakukan bahwa menggunakan satu buah hidden layer dengan atribut sebanyak 56 buah dan menghasilkan tiga buah output yang terdiri dari class 1, class 2 dan class 3 yaitu class kanker paru secara patologis. Pada class 1 terdiri dari 9 observasi, class 2 terdiri dari 13 observasi dan class 3 terdiri dari 10 observasi pada jumlah sample sebanyak 32 orang dan jumlah atribut sebanyak 57 buah, satu buah atribut class label dan 56 atribut berupa nominal integer dengan batasan antara 0-3 class pada UCI Dataset dengan class 1 sebagai kanker terbanyak yaitu sebesar 85% dengan nilai akurasi sebesar 75% menggunakan algoritma neural network.

IV. KESIMPULAN

Penelitian ini dilakukan dengan menggunakan data set berupa kanker paru dengan jumlah contoh sebanyak 32 pasien dan jumlah atribut sebanyak 57 buah dengan satu buah berupa class label dan 56 buah berupa atribut. Penelitian kali ini menggunakan Neural Network dengan nilai akurasi sebesar 75%. Dibutuhkan penelitian lebih lanjut untuk meningkatkan nilai akurasi

V. REFERENSI

- Aeberhard, S. (1991). *UCI Dataset*. Retrieved from <https://archive.ics.uci.edu/ml/datasets/Lung+Cancer>
- AG Farizawani, M. P. (2019). A review of artificial neural network learning rule based on multiple variant of conjugate gradient approaches . *Journal of Physics: Conference Series*, 1-13.
- Elizabeth, M. (2021). *Mount Elizabeth*. Retrieved from <https://www.mountelizabeth.com.sg/id/facilities-services/centre-excellence/cancer/lung-cancer>
- Ichsan. (2013, November). Sistem Pendukung Keputusan Pemilihan Penerima Beasiswa Mahasiswa Kurang Mampu Pada STMIK BUDIDARMA Medan Menerapkan Metode Profile Matching. *Kursor*, 5(1), 2. Retrieved April 14, 2016, from <http://pelita-informatika.com/berkas/jurnal/1.%20TM%20Syahru.pdf>
- Napitupulu, D. (2014). Studi Validitas Dan Realibilitas Faktor Sukses Implementasi E-Government Berdasarkan Pendekatan Kappa. *Journal of Information Systems*, 72.
- Nicholson, C. (n.d.). *Pathmind*. Retrieved from <https://wiki.pathmind.com/neural-network>
- RapidMiner. (2020). *rapidminer*. Retrieved from [rapidminer: https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/neural_nets/neural_net.html](https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/neural_nets/neural_net.html)
- S. Aeberhard, S. (1991). *UCI Dataset Retrieved*. Retrieved from <https://archive.ics.uci.edu/ml/datasets/Lung+Cancer>.
- Willy, d. T. (2019, 06 21). *alodokter.com*. Retrieved from <https://www.alodokter.com/kanker-paru-paru>
- Xiongwen Pang, Y. Z. (2018). An innovative neural network approach for stock market prediction. *Springer Science+Business Media*.