

PENGARUH TRANSFORMASI DATA PADA METODE *LEARNING VECTOR QUANTIZATION* TERHADAP AKURASI KLASIFIKASI DIAGNOSIS PENYAKIT JANTUNG

Arafa Rahman Aziz^{1*}, Budi Warsito², Alan Prahutama³

^{1,2,3} Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

e-mail: arafa27aziz@students.undip.ac.id

ABSTRACT

Learning Vector Quantization (LVQ) is a type of Artificial Neural Network with a supervised learning process based on competitive learning. Despite the absence of assumptions in LVQ is an advantage, it can be a problem when the predictor variables have big different ranges. This problems can be overcome by equalizing the range of all variables by data transformation so that all variables have relatively same effect. Heart Disease UCI dataset which used in this study is transformed by several transformation methods, such as minmax, decimal scaling, z-score, mean-MAD, sigmoid, and softmax. The result show that the six transformed data can provide better LVQ classification accuracy than the raw data which has 75.99% for training performance accuracy. LVQ classification accuracy with data transformation of minmax, decimal scaling, z-score, mean-MAD, sigmoid, and softmax are 89.16%, 88.22%, 89.7%, 90.1%, 88.17% and 92.18%. Based on the One-way ANOVA test and DMRT post hoc test known that there are significant differences between the results of the classification with data transformations and raw data in 0,05 significant level of α . It is also known that the best data transformation methods are softmax for training and sigmoid for testing.

Keywords: heart disease, neural network, learning vector quantization, classification, data transformation

1. PENDAHULUAN

Jaringan Syaraf Tiruan (JST) adalah suatu sistem yang memiliki kemiripan dengan jaringan syaraf makhluk hidup dalam memproses informasi. Obyek pengamatan dalam JST dikenali dengan melakukan suatu pelatihan yang terstruktur dan terus menerus sampai sistem jaringan tersebut mampu mengenali obyek tersebut. JST dapat diterapkan untuk menyelesaikan permasalahan klasifikasi tanpa adanya berbagai asumsi klasik. *Learning Vector Quantization* (LVQ) merupakan salah satu jenis JST dengan proses pembelajaran terawasi yang berbasis *competitive learning* dimana hanya neuron pemenang saja yang diperhatikan. Neuron pemenang ditentukan dengan menghitung jarak *euclid* yang paling kecil antara vektor masukan dengan bobot neuron.

Permasalahan muncul ketika variabel-variabel prediktor memiliki jangkauan data yang saling berbeda dengan perbedaan yang cukup besar. Jangkauan yang besar memungkinkan variabel tersebut memiliki selisih yang besar pada perhitungan jarak *euclid* yang berakibat variabel dengan jangkauan kecil kehilangan pengaruh dalam menentukan hasil klasifikasi. Solusi permasalahan ini yaitu dengan menyamakan atau mendekatkan jangkauan antar variabel sehingga menjadi relatif sama dimana data ditransformasikan dengan cara tertentu. Proses ini dilakukan sebelum proses pengolahan data.

Penelitian klasifikasi dengan metode LVQ untuk data *Heart Disease* UCI dengan variabel 10 pilihan oleh Hidayati (2012) dilakukan tanpa melakukan transformasi data terlebih dahulu menghasilkan akurasi pelatihan 66,79% [5]. Data *Heart Disease* UCI tersebut terdiri dari variabel kategorik dan numerik. Variabel numerik yang digunakan terdiri dari faktor usia, tingkat kolesterol, tekanan darah yang mana memiliki jangkauan yang saling berbeda. Sehingga peneliti menduga bahwa tidak maksimalnya akurasi tersebut karena adanya perbedaan jangkauan data yang cukup besar.

Penelitian mengenai pengaruh transformasi data terhadap akurasi klasifikasi *neural network* pernah dilakukan oleh Chamidah *et al* (2012) dan Easa dan Arabo (2017) menggunakan model *Backpropagation*. Beberapa macam transformasi data yang berbeda diuji pada kedua penelitian tersebut seperti *minmax*, *decimal scaling*, *z-score*, *mean-MAD*, *median-MAD*, *softmax*, *statistical column*, dan *norm transformation*. Pada penelitian Chamidah *et al* (2012) metode *minmax* menjadi metode paling baik sedangkan pada penelitian Easa dan Arabo (2017) metode *Mean-MAD* menjadi metode terbaik diantara beberapa metode lain setelah diuji dengan beberapa *dataset* dari UCI.

Pada penelitian ini dilakukan klasifikasi dengan metode LVQ menggunakan enam metode transformasi data yang sesuai dengan kebutuhan klasifikasi yaitu *minmax*, *decimal scaling*, *z-score*, *mean-MAD*, *sigmoid*, dan *softmax* termasuk juga klasifikasi dengan data asli. Adapun data yang digunakan adalah data *Heart Disease* UCI dengan 13 variabel prediktor. Penelitian ini bertujuan untuk mengetahui pengaruh transformasi data terhadap akurasi klasifikasi data *Heart Disease* UCI baik pada proses pelatihan maupun pengujian.

2. TINJAUAN PUSTAKA

2.1. Penyakit Jantung Koroner

Istilah penyakit jantung (*heart disease*) sering digunakan untuk menerangkan sebuah penyakit pembuluh darah jantung [8]. Pembuluh darah tersebut disebut arteri koroner yang berfungsi untuk menyuplai oksigen dan berbagai zat gizi untuk otot jantung. Gangguan pada pembuluh darah ini biasa disebut dengan Penyakit Jantung Koroner (PJK) atau sering disebut juga penyakit arteri koroner (*Coronary Artery Disease*). Gangguan ini disebabkan oleh penumpukan kolesterol dan bahan lainnya yang disebut plak di dinding bagian dalam arteri yang mengakibatkan berkurangnya aliran darah melalui arteri. Dampaknya otot jantung tidak bisa mendapatkan darah atau oksigen yang sesuai kebutuhan.

2.2. Pre-Processing

Pre-Processing adalah langkah-langkah yang dilakukan sebelum tahap pelatihan, agar data sesuai dengan metode analisis yang digunakan. Salah satu tahap *pre-processing* adalah mengestimasi data yang hilang. Data yang hilang secara sederhana dapat diestimasi menggunakan nilai rata-rata atau median untuk data bilangan riil, dan nilai modus untuk data bilangan bulat.

Tahapan *pre-processing* berikutnya adalah transformasi data. Data yang memiliki beberapa variabel dengan jangkauan nilai yang berbeda-beda mengakibatkan variabel dengan nilai atau jangkauan yang besar mempunyai pengaruh yang lebih kuat dalam fungsi biaya daripada variabel dengan nilai atau jangkauan yang kecil [9]. Permasalahan ini dapat diatasi dengan dilakukan penyekalaan ulang terhadap data agar dimiliki jangkauan yang serupa pada semua variabel. Hal ini berlaku untuk permasalahan klasifikasi. Adapun metode transformasi yang digunakan antara lain:

- **Minmax**

Metode ini mengganti data dari suatu jangkauan ke jangkauan baru yang lain dengan tetap menjaga hubungan data yang asli [10]. Metode ini umumnya digunakan untuk menskalakan ulang data pada jangkauan 0 s.d. 1 sehingga rumusnya menjadi:

$$x'_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (1)$$

- *Decimal Scaling*

Metode decimal scaling bekerja dengan cara menggeser titik desimal pada nilai data. Banyaknya penggeseran titik desimal bergantung pada nilai maksimum absolut dari masing-masing fitur data atau variabel [10]. Rumus transformasi *decimal scaling* dengan j adalah nilai terkecil sedemikian sehingga nilai $\max(|x'_i|) < 1$ sebagai berikut:

$$x'_i = \frac{x_i}{10^j} \quad (2)$$

- *Z-Score*

Z-score atau disebut juga skor baku adalah angka yang menunjukkan seberapa jauh skor mentah dari skor rata-rata (μ) dalam satuan simpangan baku (σ). [12]. Metode ini mampu mereduksi pengaruh *outlier* terhadap sebaran data hasil transformasi. Adapun rumusnya sebagai berikut:

$$x' = \frac{x - \mu}{\sigma} \quad (3)$$

- *Mean-MAD*

Metode ini mirip dengan metode z-score hanya saja nilai simpangan baku diganti dengan nilai *Mean Absolute Deviation*. Adapun rumusnya sebagai berikut [3]:

$$x'_i = \frac{x_i - \text{mean}(x)}{MAD} \quad (10)$$

Adapun nilai MAD ditentukan dengan:

$$MAD = \frac{1}{n} \sum_{i=1}^{10} |x_i - \text{mean}(x)| \quad (11)$$

- *Sigmoid*

Fungsi sigmoid merupakan fungsi yang sering digunakan dalam transformasi data sebagai fungsi aktivasi dalam JST dimana ia merubah data menjadi bernilai minimal 0 dan maksimal 1. Rumus fungsi sigmoid yang digunakan sebagai berikut [2]:

$$f(x) = \frac{1}{1+e^{-x}} \quad \text{untuk } -\infty < x < \infty \quad (3)$$

- *Softmax*

Metode *softmax* mentransformasi data menggunakan fungsi sigmoid dengan memanfaatkan nilai skor baku sebagai *input*-nya. Adapun rumusnya sebagai berikut [2]:

$$x' = \frac{1}{1 + e^{-\frac{x-\mu}{\sigma}}} \quad (12)$$

2.3. Learning Vector Quantization

Learning Vector Quantization (LVQ) adalah suatu metode untuk melakukan pelatihan terhadap lapisan-lapisan kompetitif yang terawasi. Lapisan kompetitif akan belajar secara otomatis untuk melakukan klasifikasi terhadap vektor input yang diberikan. Apabila vektor input memiliki jarak yang sangat berdekatan, maka vektor-vektor tersebut dikelompokkan ke dalam kelas yang sama [15].

LVQ merupakan salah satu jenis ANN yang berbasis *competitive learning*, dimana dari nilai keluaran yang diberikan neuron dalam *layer* keluaran hanya neuron pemenang

(neuron yang memiliki nilai terkecil) saja yang diperhatikan. Neuron tersebut akan mengalami pembaharuan bobot. Pembaharuan bobot pada neuron pemenang bisa dilakukan dengan penambahan maupun pengurangan [9].

Parameter-parameter yang digunakan dalam LVQ yang pertama adalah laju pembelajaran (*Learning Rate*) yang disimbolkan dengan α (alfa) memiliki nilai $0 < \alpha < 1$. Jika nilai α terlalu besar maka nilainya menjadi tidak stabil, tetapi jika terlalu kecil akan membuat proses pembelajaran menjadi lebih lama. Parameter yang kedua adalah *Decrease of Alfa* (*Dec α*) yaitu penurunan laju pembelajaran. Nilai laju pembelajaran akan terus diturunkan setiap kali kenaikan iterasi dengan mengalikannya dengan *Dec α* .

Algoritma klasifikasi LVQ dipaparkan sebagai berikut:

1. Menentukan bobot awal, jumlah neuron, laju pembelajaran (α), dan jumlah *epoch* maksimal. Bobot awal dicari dengan memberikan nilai tengah dari jangkauan dari nilai setiap fitur.
2. Hitung kuantisasi atau jarak vektor (d) ke semua neuron menggunakan rumus

$$d_i = \sqrt{\sum_{i=1}^r (x_i - w_i)^2} \quad (13)$$

Neuron dengan kuantisasi terkecil dipilih untuk dijadikan pemenang.

3. Perbaharui bobot neuron pemenang menggunakan rumus

$$w_i(\text{baru}) = \begin{cases} w_i(\text{lama}) + \alpha(x_i - w_i(\text{lama})), & \text{jika } \omega_{wj} = \omega_x \\ w_i(\text{lama}) - \alpha(x_i - w_i(\text{lama})), & \text{jika } \omega_{wj} \neq \omega_x \end{cases} \quad (14)$$

4. Kembali ke langkah 2 hingga semua vektor memiliki kelas.
5. Perbaharui laju pembelajaran (α) dengan rumus

$$\alpha_{\text{baru}} = (\text{Dec } \alpha) \times \alpha_{\text{lama}} \quad (15)$$

6. Uji kondisi berhenti ! Kondisi berhenti ditentukan dengan dicapainya nilai toleransi kesalahan (*error*) atau jika telah mencapai *epoch* maksimal. Jika kondisi belum terpenuhi ulangi langkah 2.

6.1. Pengukuran Ketepatan Klasifikasi

Pengukuran kinerja dalam klasifikasi dapat dilihat berdasarkan nilai *Apparent Error Rate* (APER). APER didefinisikan sebagai hasil observasi pada proses pelatihan yang mengalami kesalahan klasifikasi oleh fungsi klasifikasi [6]. Nilai APER dapat dicari dengan melihat hasil klasifikasi yang disajikan oleh matriks konfusi.

Tabel 1. Matriks Konfusi

Kelas Asli Hasil Observasi (i)	Kelas Hasil Prediksi (j)	
	Kelas = 1	Kelas = 2
Kelas = 1	<i>True Positive</i>	<i>False Positive</i>
Kelas = 2	<i>False Negative</i>	<i>True Negative</i>

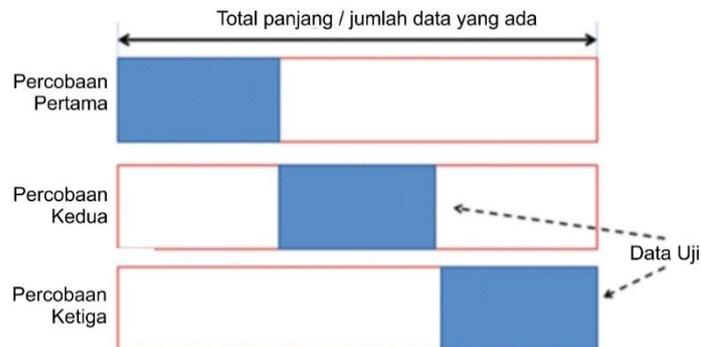
Nilai APER dan akurasi dihitung menggunakan rumus:

$$APER = \frac{FP + FN}{TP + TN + FP + FN} \quad (16)$$

$$Akurasi = (1 - APER) \times 100\% \quad (17)$$

6.2. Validasi Data

Metode *K-Fold Cross Validation* bekerja dengan menggunakan pendekatan bahwa setiap data digunakan dalam jumlah yang sama untuk pelatihan dan tepat satu kali untuk pengujian [9]. Ilustrasi validasi K Fold dengan tiga kali percobaan ditunjukkan oleh Gambar 1.



Gambar 1. Ilustrasi pembagian data pada evaluasi metode *K-Fold Cross Validation*

6.3. ANOVA Satu Arah

Uji ANOVA (Analisis of Variance) satu arah pada penelitian ini bertujuan untuk mengetahui apakah perlakuan atau faktor yang diberikan dari suatu percobaan memberikan perbedaan yang nyata (signifikan). Adapun penyebutan satu arah dikarenakan hanya satu faktor (yaitu perlakuan) yang diselidiki [16]. Pengambilan kesimpulannya didasarkan pada nilai F_{hitung} yang dibandingkan dengan nilai F_{tabel} yang terdapat pada tabel ANOVA dimana nilai F_{tabel} diperoleh dari tabel distribusi F. Jika nantinya H_0 ditolak atau H_1 diterima maka selanjutnya dapat dilakukan uji perbandingan berganda (uji lanjut). Uji lanjut digunakan untuk menentukan perlakuan mana yang menyebabkan H_0 ditolak yang salah satunya adalah uji Duncan Multiple Range Test (DMRT).

Tabel 2. ANOVA Satu Arah

Sumber Keragaman	Derajat Bebas	Jumlah Kuadrat	Kuadrat Tengah	F_{hitung}	F_{tabel}
Perlakuan	a-1	JKP	KTP	KTP/KTG	$F_{a-1;a(n-1)}(\alpha)$
Galat	a(n-1)	JKG	KTG		
Total	a.n-1	JKT			

7. METODE PENELITIAN

7.1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari lokasi Cleveland Clinic Foundation. Data dengan judul "*HeartDisease UCI*" diunduh pada tanggal 1 Februari 2020 diperoleh dari website ICS UCI (<https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data>) dengan ukuran data 303.

7.2. Variabel Penelitian

Data yang akan dianalisis merupakan data *HeartDisease UCI* Cleveland yang terdiri dari 14 variabel hasil reduksi dari 76 variabel. Data tersebut merupakan data diagnosis penyakit jantung dengan 13 variabel prediktor. Adapun variabel-variabel yang digunakan sebagai berikut:

Tabel 3. Variabel Penelitian Heart Disease UCI Cleveland

Variabel	Deskripsi	Keterangan
X ₁	Umur (<i>age</i>)	Numerik
X ₂	Jenis kelamin (<i>sex</i>)	1: Laki-Laki 0: Perempuan
X ₃	Tipe sakit dada (<i>cp</i>)	1: Typical angina 2: Atypical angina 3: Non-anginal pain 4: Asymptomatic
X ₄	Tekanan darah sistol (<i>trestbps</i>)	Numerik
X ₅	Kadar kolesterol (<i>chol</i>)	Numerik
X ₆	Gula darah puasa (<i>fbs</i>)	0: <120 mg/dl 1: >120 mg/dl
X ₇	Hasil elektrokardiografik saat istirahat (<i>restecg</i>)	0: Normal 1: Adanya kelainan gelombang ST-T 2: Hipertrofil ventrikel
X ₈	Maksimum denyut jantung yang dapat dicapai (<i>thalac</i>)	Numerik
X ₉	Ada tidaknya sakit dada akibat latihan yang diberikan (<i>exang</i>)	1: Ya 0: Tidak
X ₁₀	Puncak segmen ST yang diperoleh dari latihan relatif terhadap istirahat (<i>oldpeak</i>)	Numerik
X ₁₁	Kemiringan segmen ST untuk latihan maksimal (<i>slope</i>)	1: <i>Up sloping</i> 2: <i>Horizontal sloping</i> 3: <i>Down sloping</i>
X ₁₂	Pembuluh darah utama yang diwarnai oleh fluros kopi (<i>ca</i>)	0, 1, 2, 3
X ₁₃	Tipe kerusakan pembuluh darah (<i>thal</i>)	1: Normal 2: Cacat tetap 3: Cacat reversible
Y	Diagnosis Penyakit Jantung	1: High risk 2: Low risk

7.3. Tahapan Analisis

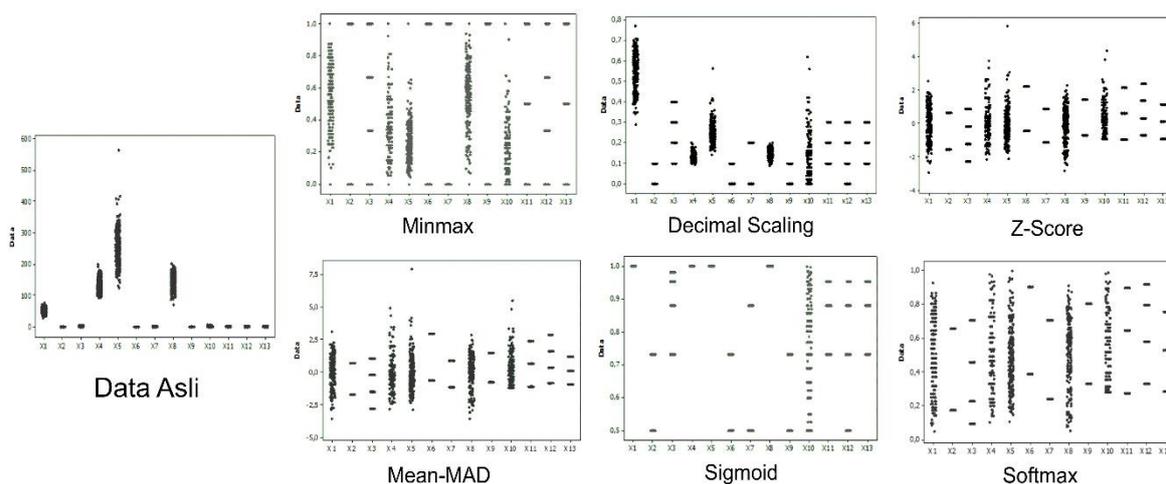
Program komputer yang digunakan untuk mendukung proses penelitian ini adalah *software* Microsoft Excel 2013, Minitab 14, dan Matlab 2015a. Adapun langkah-langkah pengolahan data yang dilakukan sebagai berikut:

1. Menyiapkan data *Heart Disease* UCI Cleveland.
2. Melakukan proses *pre-processing* berupa estimasi data hilang menggunakan *software* Microsoft Excel 2013.
3. Membagi data menjadi dua bagian yaitu data latih dan data uji. Proporsi yang digunakan adalah 67:33 (202:101) dan 80:20 (240:63)
4. Melakukan *pre-processing* berupa transformasi untuk data latih menggunakan transformasi yang sudah ditentukan, yaitu: *minmax*, *mean-MAD*, *z-score*, *decimal scaling*, *sigmoid*, dan *softmax* menggunakan *software* Microsoft Excel 2013.
5. Melakukan transformasi untuk data uji dengan menyesuaikan parameter transformasi pada data latih.
6. Melakukan *input* data latih dan data uji tanpa transformasi pada *software* Matlab 2015a.
7. Menentukan jumlah neuron, dan parameter. Jumlah neuron yang akan diuji adalah 5, 15, dan 25. Parameter *learning rate* yang diujikan yaitu: 0,25; 0,125; dan 0,025. Sedangkan untuk parameter lainnya nilainya tetap yaitu maksimum *epoch* yang digunakan adalah 300, Dec alfa yang digunakan adalah 0,1, dan nilai Min alfa yang digunakan adalah 0,001.

8. Melakukan proses pelatihan dan pengujian metode *Learning Vector Quantization*. Ulangi hingga 10 kali dan catat tiap akurasi dan peformma yang dihasilkan. Uji pula validasinya menggunakan *K-Fold Validation*.
9. Melakukan (langkah 8) minimal sekali menggunakan peformma terbaik sebagai kriteria pemberhentian iterasi pada *syntax* 'net.trainParam.goal' pada program Matlab 2015a agar didapat akurasi dan bobot terbaik. Cetak pula bobot akhir yang dihasilkan.
10. Pilih parameter jumlah neuron, *learning rate*, dan *maximum epoch* yang menghasilkan akurasi terbaik pada (langkah 8) sebagai *input* untuk proses LVQ dengan data tertransformasi.
11. Melakukan (langkah 8) menggunakan data tertransformasi pada (langkah 4) dan parameter pada (langkah 10).
12. Melakukan (langkah 9) untuk klasifikasi LVQ dengan data tertransformasi agar diperoleh hasil optimal.
13. Membandingkan rata-rata akurasi klasifikasi metode LVQ dengan data tanpa transformasi dengan semua data tertransformasi. Bandingkan juga akurasi terbaik yang dihasilkan.
14. Melakukan pengujian secara statistik untuk memperoleh kesimpulan yang objektif apakah penggunaan keenam metode transformasi memberikan pengaruh yang nyata menggunakan uji ANOVA Satu Arah dan uji lanjut DMRT.

8. HASIL DAN PEMBAHASAN

Data *Heart Disease* UCI terdiri dari 13 variabel prediktor. Dua diantaranya terdapat data hilang. Data hilang terdapat pada variabel X_{12} sebanyak tiga dan X_{13} sebanyak dua. Kedua variabel tersebut merupakan variabel dengan tipe kategorik sehingga penggantian nilai data hilang dilakukan menggunakan nilai modus pada masing-masing kelasnya.



Gambar 2. Plot Data *Heart Disease* UCI

Berdasarkan Gambar 2 ditampakkan bahwa pada data asli persebaran data dari beberapa variabelnya tidak merata yang mengakibatkan hasil klasifikasi akan banyak dipengaruhi oleh variabel yang memiliki jangkauan paling besar yaitu variabel X_5 , diikuti variabel X_4 dan X_8 . Pada data yang telah ditransformasi penyebaran data menjadi lebih merata meskipun tidak sempurna. Masing-masing transformasi memiliki penyebaran data yang berbeda. Pada transformasi minmax nilai minimal danmaksimal data benar-benar sama. Pada transformasi *decimal scaling* jangkauan data masih cukup merata meskipun nilainya berbeda. Pada transformasi *z-score* dan *mean-MAD* sebagian besar data berada

pada jangkauan yang seimbang dengan tidak memaksakan *outlier* masuk kedalam jangkauan secara umumnya. Pada transformasi *softmax* data menyebar cukup merata.

Perhitungan pertama yang dilakukan adalah klasifikasi LVQ dengan data asli yang dilakukan sesuai dengan algoritma yang sudah dibahas sebelumnya. Hasilnya akurasi terbaik diperoleh pada proporsi data latih 67% : data uji 33% , laju pembelajaran $\alpha = 0,125$, dan jumlah neuron 25. Adapun akurasi terbaik yang diperoleh adalah 75,99% untuk rata-rata pelatihan dari 10 kali percobaan dengan 61,78% untuk rata-rata pengujian dan 80,22% untuk akurasi pelatihan maksimal yang dapat dicapai.

Parameter terbaik yang sudah diperoleh kemudian diaplikasikan pada klasifikasi LVQ dengan transformasi data. Hasilnya diperoleh metode transformasi yang mampu menghasilkan akurasi tertinggi untuk proses pelatihan yaitu *softmax*, sedangkan untuk proses pengujian akurasi tertinggi dihasilkan oleh metode sigmoid. Hasil selengkapnya dapat dilihat pada tabel berikut

Tabel 4. Perbandingan Akurasi Klasifikasi LVQ dengan Transformasi Data

Jenis Data	Nilai Rata-Rata (%)		Akurasi Maksimum* (%)	K-Fold Cross Validation (%)			
	Pelatihan	Pengujian		F1	F2	F3	s
Asli	75,99	61,78	81,19	75,59 (60,5)	77,77 (63,66)	75,99 (61,78)	0,95 (1,3)
Minmax	89,16	74,95	91,58	87,03 (78,71)	86,98 (79,7)	89,16 (74,95)	1,02 (2,05)
Dec. Scaling	88,22	75,54	92,08	86,83 (82,48)	87,92 (80,5)	88,21 (75,54)	0,6 (2,91)
Z-Score	89,70	77,13	93,07	89,06 (81,29)	88,91 (79,21)	89,7 (77,13)	0,34 (1,7)
Mean MAD	90,10	78,42	94,06	88,42 (81,29)	87,62 (78,32)	90,1 (78,22)	1,03 (1,08)
Sigmoid	88,17	81,09	92,08	88,22 (81,49)	89,75 (80,01)	88,17 (81,09)	0,74 (0,58)
Softmax	92,18	76,63	96,04	90,00 (82,57)	90,69 (80,10)	92,18 (76,63)	0,3 (2,4)

Catatan: *menunjukkan hasil akurasi klasifikasi dengan data pelatihan, tanda kurung “()” menyatakan hasil pengujian

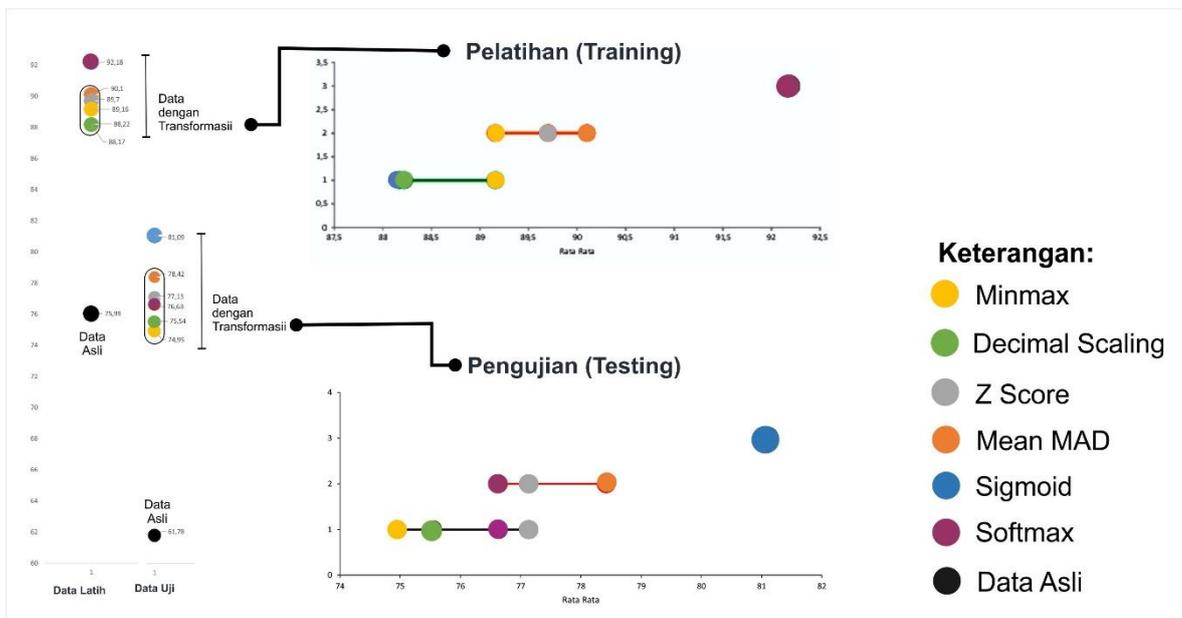
Berdasarkan Tabel 15 dapat diketahui bahwa terdapat perbedaan yang cukup jelas yaitu lebih dari 10% antara klasifikasi LVQ tanpa transformasi dengan klasifikasi LVQ dengan transformasi data baik pada proses pelatihan maupun pengujian. Pada percobaan yang diulang sepuluh kali didapat metode transformasi softmax mampu menghasilkan akurasi data pelatihan paling tinggi sebesar 92,18% dengan akurasi 76,63% untuk proses pengujian. Metode transformasi data yang memberikan hasil terendah pada proses pelatihan adalah sigmoid dengan akurasi 88,17% namun mampu memberikan hasil paling baik dalam proses pengujian dengan akurasi sebesar 81,09%. Hasil validasi dengan *K-Fold Cross Validation* menunjukkan hasil yang cukup stabil karena nilai standar deviasi (s) yang masih cukup kecil kurang dari 3.

Hasil rata-rata akurasi pada Tabel 2 dilakukan pengujian secara statistik dengan ANOVA Satu Arah untuk mengetahui apakah perbedaan akurasi tersebut signifikan. Adapun nilai α yang digunakan adalah 5%. Perlakuannya adalah jenis transformasi data.

Tabel 5. Tabel ANOVA Pelatihan dan Pengujian

Data	Tabel ANOVA					
	Sumber Keragaman	Db	JK	KT	F _{hitung}	F _{tabel}
Pelatihan	Transformasi Data	6	1695,366	282,561	145,535	2,25
	Galat	63	122,317	1,942		
	Total	69	1817,682			
Pengujian	Sumber Keragaman	Db	JK	KT	F _{hitung}	F _{tabel}
	Transformasi Data	6	2309,152	384,859	59,396	2,25
	Galat	63	408,213	6,48		
	Total	69	2717,366			

Tabel ANOVA di atas menunjukkan bahwa nilai $F_{hitung} > F_{tabel}$ baik pada hasil pelatihan maupun pengujian. Sehingga H_0 ditolak dan dapat disimpulkan bahwa adanya pengaruh transformasi data terhadap akurasi klasifikasi LVQ dengan data *Heart Disease* UCI.



Gambar 3. Hasil Uji DMRT

Hasil uji lanjut menunjukkan bahwa keenam metode transformasi yang diujikan berbeda secara signifikan terhadap akurasi dengan data asli. Lebih lanjut diketahui bahwa metode transformasi data yang paling baik adalah softmax untuk pelatihan dan sigmoid untuk pengujian.

9. KESIMPULAN

Berdasarkan hasil dan pembahasan, dapat diperoleh kesimpulan sebagai berikut:

1. Hasil klasifikasi LVQ data *Heart Disease* UCI menggunakan enam metode transformasi yaitu minmax, decimal scaling, z score, mean MAD, sigmoid, dan z score, keseluruhannya lebih baik secara signifikan daripada klasifikasi dengan tanpa transformasi data baik pada pelatihan maupun pada pengujian berdasarkan uji ANOVA satu arah pada taraf signifikansi $\alpha = 5\%$ dengan nilai $F_{hitung} = 145,535 > F_{tabel} = 2,25$ untuk pelatihan dan nilai $F_{hitung} = 59,396 > F_{tabel} = 2,25$ untuk pengujian.

2. Transformasi data yang menghasilkan akurasi klasifikasi LVQ data Heart Disease UCI terbaik berdasarkan uji lanjut DMRT adalah transformasi softmax untuk pelatihan dan transformasi sigmoid untuk pengujian.

DAFTAR PUSTAKA

- [1] Berry, M. W., Mohammed, A., & Yap, B.W. 2016. *Soft Computing in Data Science*. Singapura: Springer
- [2] Chamidah, N., Winarto, & Salamah, U. 2012. *Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Descent Adaptive Gain (BPGDAG) untuk Klasifikasi*. ITSMART, Vol 1, No 1: hlm 28-33
- [3] Eesa, A., & Arabo, W. 2017. *A Normalization Methods for Backpropagation: A Comparative Study*. Science Journal of University of Zakho Vol 5, No 4: hlm 319-323
- [4] Hendrawati, T. 2015. *Kajian Metode Imputasi dalam Menangani Missing Data*. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS 2015. No 67, hlm: 637-642
- [5] Hidayati, N., & Warsito, B. 2012. *Prediksi Terjangkitnya Penyakit Jantung dengan Metode Learning Vector Quantization*. Media Statistika, Vol. 3, No. 1: hlm 21-30
- [6] Johnson, R. & Wichern, D. 2007. *Applied Multivariate Statistical Analysis*. Edisi 6. USA: Pearson Education Inc
- [7] Kusumadewi, S. 2004. *Membangun Jaringan Syaraf Tiruan (Menggunakan Matlab dan Excel Link)*. Edisi 1. Yogyakarta: Graha Ilmu
- [8] Lovastatin, K. 2006. *Penyakit Jantung dan Penanganannya dengan Metode Alami*. Diterjemahkan oleh: Slamet Rianto. Jakarta: PT. Pustaka restasi Pustakaraya. Terjemahan dari: Mayo Clinic On Heart Disease and High Blood Pressure
- [9] Prasetyo, E. 2014. *Data Mining - Mengolah Data menjadi Informasi Menggunakan Matlab*. Yogyakarta: ANDI
- [10] Pan, J., Zhuang, Y., & Fong, S. 2016. *The Impact of Data Normalization on Stock Market Prediction: Using SVM and Technical Indicators*. CCIS, Vol 652, No 7: hlm 72-88
- [11] Siang, J. 2005. *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab*. Yogyakarta: ANDI.
- [12] Sufren, N. Y. 2013. *Mahir Menggunakan SPSS Secara Otodidak*. Jakarta: PT Elex Media Komputindo
- [13] Soeharto, I. 2004. *Penyakit Jantung Koroner dan Serangan Jantung*. Jakarta: PT Gramedia Pustaka Utama
- [14] [UCI] University of California Irvine. 1988. *Heart Disease Dataset*. <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>. Diakses: 9 Februari 2020
- [15] Warsito, B. 2009. *Kapita Selekta Statistika Neural Network*. Semarang: BP Undip Semarang.
- [16] Widiharah, T. 2007. *Buku Ajar Rancangan Percobaan*. Semarang: Program Studi Statistika Jurusan Matematika Undip