

Klasifikasi Kemampuan Perawatan Diri Anak dengan Disabilitas Menggunakan SMOTE Berbasis *Neural Network*

Sari Susanti

Universitas BSI
e-mail: sari.srq@bsi.ac.id

Abstrak

Penyandang disabilitas merupakan kelompok minoritas terbesar didunia, dengan anak-anak menempati sepertiga dari jumlah keseluruhan penyandang disabilitas. Pada penerapannya proses diagnosis dan klasifikasi dimensi disabilitas membutuhkan ahli terapis okupasi. Jumlah terapis okupasi yang terbatas mengakibatkan penanganan penyandang disabilitas menjadi tertunda. Teknik data mining digunakan untuk membantu proses diagnosis yang bertujuan untuk menghindari kesalahan dalam diagnosis. Penelitian ini menggunakan dataset Scadi yang merepresentasikan masalah kemampuan perawatan diri anak dengan disabilitas. Dataset Scadi merupakan dataset baru yang belum banyak diteliti. Pada dataset Scadi terdapat permasalahan yaitu, ketidakseimbangan kelas (*imbalanced class*). Masalah tersebut menyebabkan rendahnya nilai akurasi klasifikasi. Algoritma yang diusulkan yaitu *neural network* untuk klasifikasi kemampuan perawatan diri anak dengan disabilitas, Selain *neural network* digunakan algoritma Smote untuk mengatasi masalah ketidakseimbangan kelas (*imbalanced class*) pada level data. Hasil penelitian menunjukkan bahwa metode yang diusulkan telah meningkatkan kinerja algoritma klasifikasi *neural network*, dengan meningkatkan nilai akurasi secara signifikan sebesar 90.4762 % dibandingkan hasil yang dilaporkan pada penelitian sebelumnya yaitu 83,1%.

Kata kunci: Perawatan Diri, Disabilitas, ICF-CY, *Neural Network*, *Smote*, Ketidakseimbangan Kelas.

Abstract

Persons with disabilities are the majority group in the world, with children determined one third of the total number of persons with disabilities. In its application, the process of diagnosis and classification of occupational therapist needs. The number of occupational therapists who are limited to handling persons with disabilities is delayed. Data mining techniques are used to help the diagnosis process that helps to avoid errors in diagnosis. This study uses the Scadi dataset which presents the problem of self-care ability of children with disabilities. Dataset Scadi is a new dataset that has not yet been collected. The Scadi dataset is related, that is, class imbalances (unbalanced classes). This problem causes a low value. The proposed algorithm is a neural network for the classification of self-care abilities of children with disabilities. In addition to the neural network, the Smote algorithm is used to overcome the problem of class imbalances (unbalanced levels) at the data level. The results showed that the proposed method had improved the results of neural network classification analysis, by increasing the assessment value by 90.4762% compared to the results obtained in the previous study, namely 83.1%.

Keywords: *Self-care, Disability, ICF-CY, Neural Network, Smote, Imbalanced Class.*

1. Pendahuluan

Disabilitas merupakan gangguan, keterbatasan aktivitas dan pembatasan partisipasi. Disabilitas disebut juga interaksi antara individu dengan kondisi kesehatan seperti (Cerebral palsy, sindrom Down dan depresi), faktor pribadi dan lingkungan

seperti sikap negatif (WHO, 2018). Disabilitas dapat mengganggu perkembangan alami tubuh tergantung pada jenis kelamin, usia dan lingkungan (Zarchi, Bushehri, & Dehghanizadeh, 2018). Penderita disabilitas merupakan kelompok minoritas terbesar didunia, 80% penderita

berasal dari negara-negara berkembang. Selain itu Anak-anak menempati menyandang disabilitas dengan jumlah sepertiga dari jumlah keseluruhan penyandang disabilitas di dunia (Kemenkes, 2014).

Lebih dari satu miliar orang diperkirakan hidup dengan beberapa bentuk disabilitas atau kecacatan fisik dan motorik. Data WHO menyebutkan sekitar 15% populasi di dunia antara 110 juta (2,2%) dan 190 juta (3,8%) dengan usia 15 tahun atau lebih mengalami kesulitan fungsionalitas yang signifikan (WHO, 2018). Klasifikasi dan pengkodean dimensi disabilitas pada anak-anak dan remaja merupakan proses kompleks. Proses ini membutuhkan pertimbangan keterbatasan fungsi tubuh, struktur tubuh, aktivitas dan partisipasi dalam perkembangan fisik, sosial dan psikologis (WHO, 2007).

Proses diagnosis dan klasifikasi dimensi disabilitas membutuhkan ahli terapis okupasi. Jumlah terapis okupasi yang terbatas mengakibatkan penanganan pasien disabilitas tertunda, akibatnya membuat pengobatan pasien menjadi lebih sulit dan mahal. Berdasarkan hal tersebut sistem klasifikasi masalah perawatan diri anak dengan disabilitas merupakan faktor penting dalam memilih pendekatan perawatan. Karena keragaman dan kompleksitas masalah perawatan diri dan kurangnya terapis okupasi, menggunakan sistem klasifikasi otomatis untuk masalah perawatan diri dapat membantu terapis okupasi untuk memilih pendekatan dan pengobatan terbaik (Zarchi et al., 2018).

WHO memberikan solusi untuk mempermudah klasifikasi jenis perawatan untuk penderita disabilitas dengan kerangka kerja *The International Classification of Functioning, Disability and Health for Children and Youth (ICF-CY)* (Yeh, Hou, & Chang, 2012). *ICF-CY* merupakan kerangka konseptual klasifikasi multiguna yang banyak digunakan dan merupakan versi turunan pertama dari rujukannya *ICF* (WHO, 2007). Berdasarkan hal tersebut diperlukan suatu penelitian untuk mendapatkan model yang mampu mengklasifikasikan hasil (*ICF-CY*) dengan tingkat kesalahan klasifikasi yang minimal. Penelitian sebelumnya mengenai klasifikasi kemampuan perawatan diri anak disabilitas telah dilakukan (Zarchi et al., 2018).

Masalah yang muncul pada penelitian sebelumnya yaitu, adanya ketidak-

seimbangan kelas (*imbalanced class*) sehingga tahap *pre-processing* harus dilakukan untuk meningkatkan kinerja algoritma klasifikasi (Unal & Kocer, 2013). Ketidakseimbangan kelas adalah distribusi yang tidak merata diantara kelas-kelas dimana satu kelas lebih banyak diantara kelas lain (Deepa & Punithavalli, 2013). Masalah ketidakseimbangan kelas merupakan tantangan dalam *machine learning* dan *data mining*, hal tersebut telah menarik penelitian yang signifikan beberapa tahun terakhir. Algoritma klasifikasi yang dipengaruhi oleh masalah ketidakseimbangan kelas untuk satu set data tertentu memiliki akurasi yang baik secara keseluruhan, tetapi akurasi yang sangat buruk pada kelas minoritas (Tiwari, 2014).

Secara umum terdapat dua pendekatan untuk menangani ketidakseimbangan kelas (*imbalanced class*) pada *dataset*, yaitu pendekatan level algoritma dan level data (Z.-Z. Zhang et al., 2008). Pada level algoritma, metode utamanya adalah menyesuaikan operasi algoritma yang ada untuk membuat pengklasifikasian agar lebih kondusif terhadap klasifikasi kelas minoritas (D. Zhang, Liu, Gong, & Jin, 2011). Kelemahan level algoritma jika diaplikasikan dalam algoritma klasifikasi kuat seperti *neural network* adalah waktu yang lebih lama karena adanya penyesuaian bobot dan iterasi sampai mendapat nilai yang sesuai (Korada, Kumar, & Deekshitulu, 2012). Pada level data terdapat berbagai teknik *resampling* yang digunakan untuk memperbaiki ketidakseimbangan kelas. Tiga teknik yang biasa digunakan adalah *Random Over Sampling (ROS)*, *Random Under Sampling (RUS)* dan *Synthetic Minority Over-sampling Technique (SMOTE)* yang merupakan turunan dari (ROS) (Khoshgoftaar, Gao, Napolitano, & Wald, 2014). Teknik *SMOTE* digunakan untuk mengatasi masalah *imbalanced class* pada penelitian (Putri & Wahono, 2015).

Selain penanganan masalah *imbalanced class*, pada penelitian kali ini juga dilakukan klasifikasi dengan menggunakan metode *neural network (NN)*. Metode klasifikasi dengan menggunakan metode *NN* sebelumnya telah digunakan oleh banyak peneliti untuk beberapa objek yang berbeda dari penelitian kali ini. Diantara beberapa penelitian yang menggunakan metode *NN* sebagai metode

klasifikasinya adalah penelitian yang dilakukan oleh (Zarchi et al., 2018) penelitian ini merupakan penelitian pertama untuk klasifikasi *dataset SCADI*. (Praveena & Singh, 2014) Pada penelitian ini metode klasifikasi dilakukan untuk mengklasifikasikan data gambar yang berasal dari satelit untuk kemudian dilakukan segmentasi gambar menggunakan metode *Artificial Bee Colony (ABC)* (Praveena & Singh, 2014). Selain data gambar, metode klasifikasi juga digunakan pada data set penyakit jantung coroner yang dilakukan oleh Radhimeenakshi pada tahun 2016. Pada penelitian ini dibandingkan antara metode *neural network* dengan *Support Vector Machine* terhadap data set penyakit jantung coroner *Cleveland* dan *Statlog (binomial class dan polynomial class)*. Hasil dari penelitian ini menunjukkan bahwa *neural network* menunjukkan nilai yang lebih baik dalam *Metrik Specify* dibandingkan dengan metode *support vector machine* (Radhimeenakshi, 2016).

Pada *dataset Scadi* setiap fitur adalah fitur *multidimensional* yang dibuat menggunakan kode berdasarkan kategori aktivitas perawatan diri. Pengkodean untuk setiap fitur seperti tercantum pada tabel 1 dan 2.

Tabel 1. Pengkodean Fitur

Code	Level-1	Level-2	Level-3	Level-4
ICF-CY Code =	Component Kode	10-99	0-9	0-9

Misalnya, kegiatan *toileting* seperti melakukan buang air kecil dengan tepat dikodekan dalam *ICF-CY* seperti berikut pada tabel 2

Tabel 2. Deskripsi Pengkodean Fitur

Code	Level-1	Level-2	Level-3	Level-4

Tabel 4. Penjelasan 7 kelas

Kelas	Deskripsi
1	<i>Caring for body parts problem</i>
2	<i>Toileting problem</i>
3	<i>Dressing problem</i>
4	<i>Washing oneself and Caring for body parts and Dressing problem</i>
5	<i>Washing oneself, Caring for body parts, Toileting, and Dressing problem</i>

d	d5	29	0	1
52901				

Tingkat pertama (level-1) mengklasifikasikan aktivitas umum. Tingkat ini termasuk empat kategori, yaitu Fungsi Tubuh, Struktur Tubuh, Aktivitas dan Partisipasi, dan Faktor Lingkungan. Kode Level-1 didefinisikan sebagai kombinasi angka dan huruf. Dalam *ICF-CY*, Perawatan mandiri adalah subbagian komponen Kegiatan dan Partisipasi dan dimulai dengan d5. Tingkat berikutnya (level-2, level-3, dan level-4) dikodekan oleh angka. Setiap level menjelaskan level sebelumnya dengan lebih jelas. Pada *dataset Scadi*, 29 aktivitas dipertimbangkan untuk perawatan diri berdasarkan kerangka *ICF-CY*. 7 kode dibuat untuk menggambarkan tingkat penurunan aktivitas dan komponen partisipasi. Tabel 3 menunjukkan kode batas kerusakan *Scadi* (Zarchi et al., 2018).

Masalah perawatan diri dikategorikan ke dalam tujuh kelompok oleh terapis okupasi seperti yang ditunjukkan pada Tabel 4. Kelas sasaran pada *dataset scadi* mengacu pada masalah perawatan diri yang ditentukan oleh para ahli. Setiap anak diklasifikasikan ke salah satu dari 7 kategori berdasarkan fitur perawatan diri dengan pemeriksaan terapis okupasi.

Tabel 3. Kode Gangguan *Dataset Scadi*

Code	Value	Description
xxx.0	0-4%	<i>NO impairment</i>
xxx.1	5-24%	<i>MILD impairment</i>
xxx.2	25-49%	<i>MODERATE impairment</i>
xxx.3	50-95%	<i>SEVERE impairment</i>
xxx.4	96-100%	<i>COMPLETE impairment</i>
xxx.8		<i>not specified</i>
xxx.9		<i>not applicable</i>

Dataset SCADI memiliki 7 kelas yang menjadi target atau label yang dijelaskan pada tabel 4.

Kelas	Deskripsi
6	<i>Eating, Drinking, Washing oneself, Caring for body parts, toileting, Dressing, Looking after one's health and Looking after one's safety problem</i>
7	<i>No Problem</i>

2. Metode Penelitian

Penelitian ini mengusulkan metodologi berdasarkan paradigma *data mining*. Paradigma ini mengintegrasikan algoritma *smote*, dengan algoritma klasifikasi *neural network*.

Synthetic Minority Over-sampling Technique (SMOTE)

Tujuan utama dari teknik *SMOTE* adalah untuk menangani *imbalance class*. Penggunaan teknik *Synthetic Minority Over-sampling Technique (SMOTE)* (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). menghasilkan hasil yang baik dan efektif untuk menangani *imbalance class* yang mengalami *over-fitting* pada proses teknik *over-sampling* untuk kelas minoritas (positif)(Riquelme, Ruiz, Rodriguez, & Moreno, 2008).

Smote menciptakan sebuah contoh dari kelas minoritas sintetis yang beroperasi di ruang fitur daripada ruang data. Dengan menduplikasi contoh kelas minoritas, *smote* menghasilkan contoh sintetis baru dengan melakukan ekstrapolasi sampel minoritas yang ada dengan sampel acak yang diperoleh dari nilai k tetangga terdekat. Dengan hasil sintetis pada contoh yang lebih dari kelompok minoritas, sehingga mampu memperluas area keputusan mereka untuk minoritas (Chawla et al., 2002).

Metode *smote* sebagai salah satu solusi dalam menangani data tidak seimbang dengan prinsip yang berbeda dengan Metode *oversampling*. Metode *oversampling* bekerja dengan memperbanyak jumlah data secara acak, Metode *smote* menambah jumlah data pada kelas minor sehingga setara dengan kelas mayor dengan cara membangkitkan data sintetis (Barro, Sulvianti, & Afendi, 2013).

Proses pembangkitan data sintetis yang berskala numerik berbeda dengan kategorik. Data numerik diukur jarak kedekatannya dengan jarak Euclidean sedangkan data kategorik lebih sederhana yaitu dengan nilai modus. Perhitungan jarak antar contoh kelas minor yang peubahnya

berskala kategorik dilakukan dengan rumus *Value Difference Metric (VDM)* (Cost & Salzberg, 2013) yaitu :

$$\Delta(X, Y) = w_x w_y \sum_{i=1}^N \delta(x_i, y_i)^r \quad (2.1)$$

Keterangan:

$\Delta(X, Y)$: jarak antara amatan X dengan Y

$w_x w_y$: bobot amatan (dapat diabaikan)

N : banyaknya peubah penjelas

R : bernilai 1 (jarak Manhattan) atau 2 (jarak Euclidean)

$\delta(x_i, y_i)^r$: jarak antar kategori, dengan rumus:

$$\delta(V_1, V_2) = \sum_{i=1}^N \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right|^k \quad (2.2)$$

Keterangan:

$\delta(V_1, V_2)$: jarak antara nilai V1 dan V2

C_{1i} : banyaknya V_1 yang termasuk kelas i

C_{2i} : banyaknya V_2 yang termasuk kelas i

i : banyaknya kelas; $i = 1, 2, \dots, m$

C_1 : banyaknya nilai 1 terjadi

C_2 : banyaknya nilai 2 terjadi

N : banyaknya kategori

K : konstanta (biasanya 1)

Prosedur pembangkitan data buatan untuk:

1. Data Numerik
 - a. Hitung perbedaan antar vektor utama dengan k tetangga terdekatnya.
 - b. Kalikan perbedaan dengan angka yang diacak diantara 0 dan 1.
 - c. Tambahkan perbedaan tersebut ke dalam nilai utama pada vektor utama asal sehingga diperoleh vektor utama baru.
2. Data Kategorik
 - a. Pilih mayoritas antara vektor utama yang dipertimbangkan dengan k tetangga terdekatnya untuk nilai nominal. Jika terjadi nilai samamaka pilih secara acak.
 - b. Jadikan nilai tersebut data contoh kelas buatan baru.

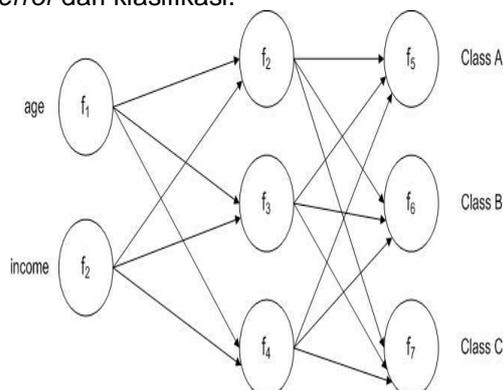
Neural network

Neural network atau jaringan saraf adalah satu set unit *input/output* yang terhubung, dimana tiap koneksinya memiliki bobot. Selama fase pembelajaran, *Neural*

network menyesuaikan bobot sehingga dapat memprediksi *class* yang benar dari *tuple* (Han & Kamber, 2006). Algoritma *Neural network* yang paling populer adalah *Backpropagation*, algoritma *backpropagation* melakukan pembelajaran pada jaringan saraf *multi layer feed forward* yang terdiri dari tiga lapisan/*layer*, yaitu: lapisan *input*, lapisan tersembunyi, dan lapisan keluaran (Han & Kamber, 2006). Pada beberapa diagram *Neural network* dimungkinkan terdapat lebih dari satu lapisan tersembunyi, meskipun kebanyakan hanya mengandung satu lapisan tersembunyi yang dirasa cukup untuk berbagai tujuan (Larose, 2006).

Backpropagation merupakan algoritma *Neural network* untuk klasifikasi yang menggunakan *gradient descent*, *backpropagation* mencari satu set bobot yang dapat memodelkan data sehingga dapat meminimalkan jarak kuadrat rata-rata antara prediksi kelas jaringan dan label kelas yang sebenarnya dari *tuple* data (Han & Kamber, 2006).

Tiap observasi data *training* diproses melalui jaringan, nilai *output* dihasilkan dari node *input*. Nilai *output* ini kemudian dibandingkan dengan nilai aktual dari variabel target dan dihitung error yang dihasilkan (Larose, 2006). *Backpropagation* melakukan proses pembelajaran secara *iteratif* yang mencoba untuk meminimalkan *error* dari klasifikasi.



Gambar 1. *Multi Layer Feed Forward Neural network* (Han & Kamber, 2006)

Proses pembelajaran dalam *backpropagation* dilakukan dengan penyesuaian bobot-bobot *neuron* dengan arah mundur berdasarkan nilai *error* dalam proses pembelajaran (Kusrini & Luthfi, 2009). Untuk mendapatkan *error* maka tahap perambatan maju (*forward propagation*) harus dikerjakan terlebih

dahulu, pada saat perambatan maju maka *neuron-neuron* diaktifkan dengan menggunakan fungsi aktivasi yang dapat dideferensiasikan, seperti fungsi sigmoid (Kusumadewi, 2004).

Dalam setiap proses, bobot relasi dalam jaringan dimodifikasi untuk meminimalkan nilai *Mean Squared Error (MSE)* antara nilai prediksi dari jaringan dengan nilai sesungguhnya. Modifikasi relasi jaringan tersebut dilakukan dalam arah mundur (Kusrini & Luthfi, 2009).

Langkah pembelajaran dalam metode *backpropagation* adalah sebagai berikut (Myatt, 2007):

1. Inisialisasi semua bobot jaringan secara acak. (Biasanya antara -1.0 hingga 1.0).
2. Untuk setiap data *training*, hitung input untuk simpul berdasarkan nilai *input* dan bobot jaringan saat itu. Dengan menggunakan rumus :

$$I_j = \sum_i w_{ij} O_i + \theta_j \quad (2.3)$$

Keterangan :

O_i = *Output* simpul i dari layer sebelumnya
 w_{ij} = bobot relasi dari simpul i pada *layer* sebelumnya ke simpul j

θ_j = bias (sebagai pembatas)

3. Berdasarkan *input* dari langkah kedua, selanjutnya membangkitkan *output* untuk simpul menggunakan fungsi aktifitas *sigmoid*

$$Output = \frac{1}{1 + e^{-input}} \quad (2.4)$$

4. Hitung nilai error antara nilai yang diprediksi dengan nilai yang sesungguhnya menggunakan rumus :

$$Error_i = output_i * (1 - Output_i) * (Target_i - Output_i) \quad (2.5)$$

Keterangan :

$Output_i$ = *Output actual* dari simpul i

$Target_i$ = nilai target yang sudah diketahui pada data *training*

5. Setelah nilai *error* dihitung, selanjutnya dibalik ke *layer* sebelumnya (*backpropagation*). Untuk menghitung nilai *error* pada *hidden layer*, menggunakan rumus.

$$Error_i = Output_i * (1 - Output_i) * \sum_{j=1}^n Error_j W_{ij} \quad (2.6)$$

Keterangan :

$Error_i$ = *Error* yang dihasilkan dari *hidden node*

$Output_i$ = Nilai *output* yang dari *hidden node*

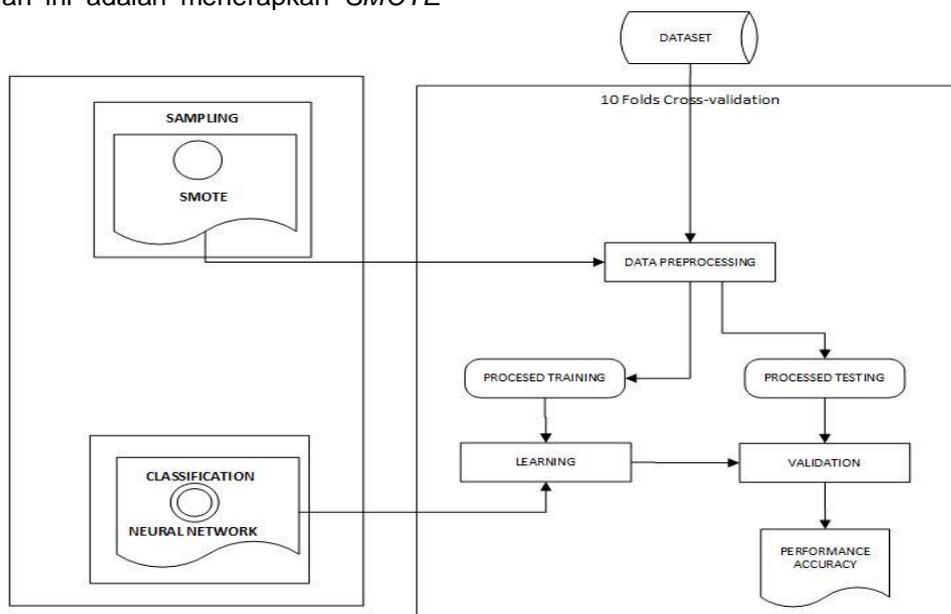
$Error_j$ = *Error* yang dihasilkan dari simpul j yang terhubung ke *output*

- W_{ij} = Bobot pada simpul i ke simpul j
6. Nilai error yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi, dengan menggunakan rumus:
- $$W_{ij} = W_{ij} + l * Error_j * Output_i \quad (2.7)$$
- Keterangan :
- W_{ij} = bobot relasi dari unit i pada layer sebelumnya ke unit j
- l = learning rate (konstantan, nilainya 0 sampai dengan 1)
- $Error_j$ = Error pada output layer simpul j
- $Output_i$ = Output dari simpul i .

Metode yang diusulkan

Metode yang diusulkan pada penelitian ini adalah menerapkan SMOTE

untuk penanganan ketidakseimbangan kelas berbasis *neural network*. Teknik SMOTE digunakan untuk memperbaiki dataset sehingga jumlah data antara satu kelas dengan kelas lainnya seimbang, selain teknik sampling pada tahap preprocessing juga diterapkan seleksi fitur menggunakan *greedy stepwise*. Dataset yang sudah diseleksi fitur kemudian diklasifikasikan menggunakan algoritma *neural network* dengan menggunakan 10-cross validation menggunakan Software WEKA v.3.8. Kemudian hasil evaluasi model akan dibandingkan dengan *confusion matrix* sebelum dilakukan SMOTE dan seleksi fitur. Metode yang diusulkan dapat dilihat pada gambar 2.



Gambar 2. Metode penelitian yang diusulkan

Model penelitian pada Gambar 2. dimulai dengan melakukan tahap preprocessing pada dataset dengan menggunakan Synthetic Minority Over Sampling (Smote) yang bertujuan untuk menambah jumlah sampel untuk kelas minoritas. Setelah data tersebut dilakukan sampling langkah selanjutnya adalah diklasifikasi dengan 10 cross-validation (iterasi 10 kali) menggunakan algoritma *Neural network*. Hasil kinerja pada model ini akan diukur dengan menggunakan akurasi.

3. Hasil dan Pembahasan

Hasil pada penelitian ini didapatkan dengan melakukan eksperimen terhadap

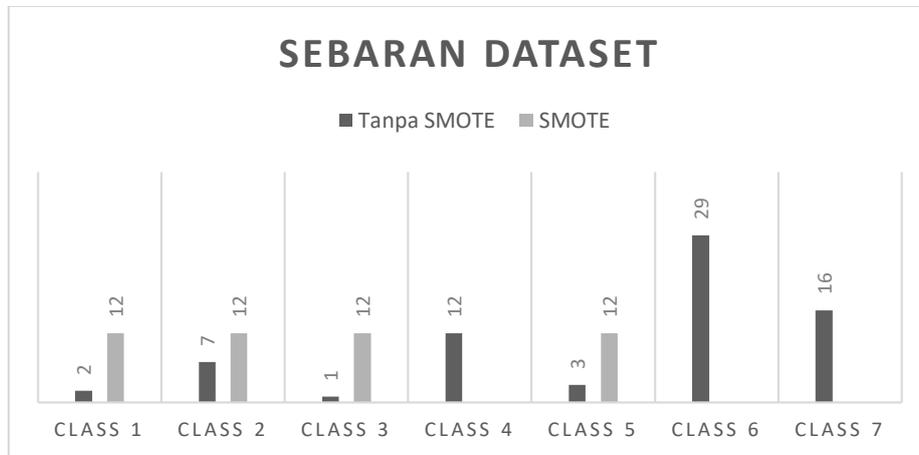
algoritma klasifikasi *data mining* yaitu *neural network* dan dua algoritma pembandingan yaitu *decision tree*, dan *logistic regression* dengan menggunakan *software weka*. Eksperimen dilakukan dengan menggunakan *Scadi datasets* yang berisi data kemampuan perawatan diri anak dengan disabilitas.

Tabel 5 Hasil Eksperimen klasifikasi

NO	Algoritma	Nilai Akurasi
1	<i>Neural network</i>	80,00%
2	<i>Decision Tree (J48)</i>	80,00%
3	<i>Logistic Regression</i>	80,00%

Tabel 5. Merupakan hasil akurasi klasifikasi kemampuan perawatan diri anak dengan disabilitas. Dimana hasil akurasi untuk ketiga algoritma yaitu *neural network*, *decision tree* dan *logistic regression* sebesar 80% yang menyimpulkan tidak ada perbedaan antara satu dengan yang lain.

Tahap selanjutnya dilakukan preprocessing menggunakan algoritma *SMOTE*. Proses ini menghasilkan jumlah data setiap kelas yang ditampilkan pada Gambar 3:



Gambar 3. Grafik Sebaran Dataset Sebelum dan Sesudah dilakukan SMOTE

Setelah dilakukan resampling kemudian dataset diklasifikasikan dengan melakukan eksperimen terhadap beberapa algoritma klasifikasi, dengan hasil yang ditampilkan pada tabel 6.

Tabel 6. Hasil Eksperimen klasifikasi dengan optimasi *Smote*

N O	Algoritma	Tanpa Optimasi	Optimasi <i>Smote</i>
1	<i>Neural network</i>	80,00%	90.4762 %
2	<i>Decision Tree (J48)</i>	80,00%	87.619 %
3	<i>Logistic Regression</i>	80,00%	84.7619 %

Tabel 6 merupakan hasil akurasi klasifikasi dengan menggunakan algoritma klasifikasi *neural network*, *decision tree* dan *logistic regression* dengan menambahkan optimasi yaitu *synthetic minority over-sampling technique (Smote)* dimana hasil akurasi mengalami peningkatan untuk semua algoritma yang diujikan. Adapun *neural network* memiliki hasil akurasi yang paling tinggi dibandingkan algoritma *decision tree* dan *logistic regression* sebesar **90.4762%**.

Berdasarkan hasil percobaan yang tertera pada Tabel 6 dilakukan analisa uji statistik uji beda *Paired Sample t-Test*, hasil yang diperoleh dicantumkan pada Tabel 7.

Tabel 7. Hasil Uji T-test Algoritma klasifikasi Tanpa Optimasi dengan Optimasi *Smote*

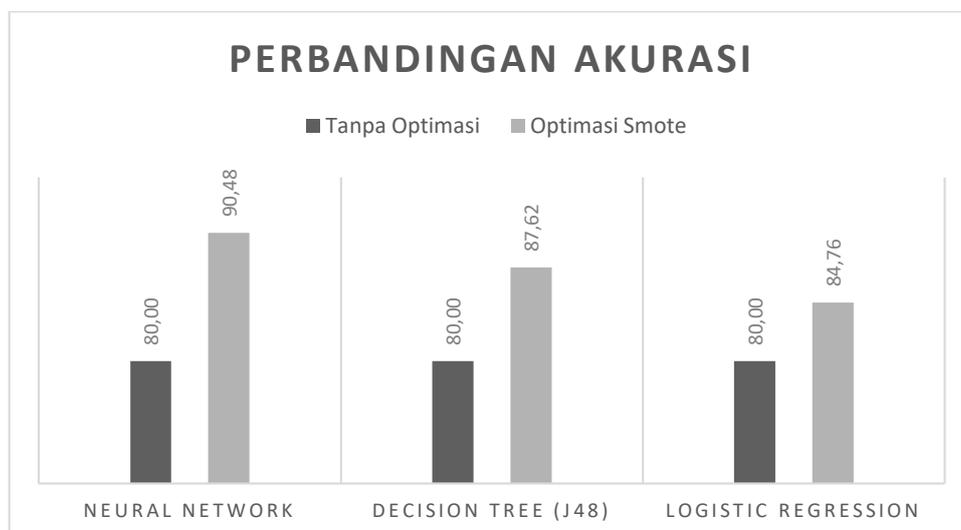
t-Test: Paired Two Sample for Means

	Tanpa Optimasi	Optimasi <i>Smote</i>
<i>Mean</i>	80	87.61903333
<i>Variance</i>	0	8.163306123
<i>Observations</i>	3	3
<i>t Stat</i>	-4.618781946	

	Tanpa Optimasi	Optimasi Smote
$P(T \leq t)$ one-tail	0.021908736	
t Critical one-tail	2.91998558	
$P(T \leq t)$ two-tail	0.043817471	
t Critical two-tail	4.30265273	

Berdasarkan hasil uji t dua sampel berpasangan pada Tabel 7 dapat diambil hipotesa berdasarkan perbandingan t hitung dan t tabel, juga berdasarkan nilai probabilitas. Nilai t hitung diwakili oleh t stat sebesar -4.618781946 dan nilai t tabel diwakili oleh t *critical two-tail* sebesar 4.30265273, maka dapat dipastikan nilai t hitung > t tabel dengan arah *negative* yang artinya H_1 diterima dan H_0 ditolak,

sedangkan diketahui nilai probabilitas sebesar 0.043817471, maka dapat dipastikan bahwa nilai probabilitas < 0,05 yang artinya terdapat perbedaan yang signifikan dari rata-rata akurasi metode (*neural network, decision tree dan logistic regression*) dengan (*neural network, decision tree dan logistic regression*) yang dikombinasikan dengan algoritma *synthetic minority over-sampling technique (smote)*.



Gambar 4. Grafik Perbandingan Performa NN dan SMOTE

Pada gambar 4. terlihat bahwa menerapkan *synthetic minority over-sampling* dapat memperbaiki performa kinerja algoritma klasifikasi dengan meningkatkan nilai akurasi.

4. Kesimpulan

Penelitian ini mengkombinasikan algoritma *synthetic minority over-sampling technique (smote)* pada algoritma *classifier neural network* untuk mengatasi masalah ketidakseimbangan kelas (*imbalanced class*) dataset *Scadi*. Berdasarkan hasil eksperimen pada penelitian ini, maka dapat ditarik kesimpulan dengan penerapan *synthetic minority oversampling technique (smote)* dapat memperbaiki kinerja

algoritma *neural network* pada dataset *Scadi*, untuk klasifikasi tujuh kelas dengan nilai akurasi sebesar **90,47%** jika dibandingkan sebelum menggunakan *smote* dengan nilai akurasi **80%**.

Referensi

- Barro, R. A., Sulvianti, I. D., & Afendi, F. M. (2013). Penerapan Synthetic Minority Oversampling Technique (Smote) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu. *Xplore*, 1(1). Retrieved from journal.ipb.ac.id/index.php/xplore/article/download/12424/9491
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE:

- Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(2), 321–357. <https://doi.org/10.1613/jair.953>
- Cost, S., & Salzberg, S. (2013). A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Deepa, T., & Punithavalli, M. (2013). An Innovative Optimization Algorithm for Feature Selection – A Comparative Study. *International Journal of Computer Science and Information Technology & Security (IJSITS)*, 3(1), 20–24.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques Second Edition*. San Francisco: Diane Cerra.
- Kemendes. (2014). *Penyandang Disabilitas Pada Anak*. Pusat Data dan Informasi Kementerian Kesehatan RI.
- Khoshgoftaar, T. M., Gao, K., Napolitano, A., & Wald, R. (2014). A comparative study of iterative and non-iterative feature selection techniques for software defect prediction. *Information Systems Frontiers*, 16(5), 801–822. <https://doi.org/10.1007/s10796-013-9430-0>
- Korada, Kumar, & Deekshitulu. (2012). Implementatioan Of Naive Bayesian Classifier and Ada-Boost Algoritm Using Maiz Expert System. *International Journal Of Information Sciences And Techniques (IJIST) Vol.2, No.3, 63-75, 2(3), 63–75*.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- Kusumadewi, S. (2004). *Membangun Jaringan Syaraf Tiruan Menggunakan Matlab & Excel Link*. Yogyakarta: Graha Ilmu.
- Larose, D. T. (2006). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Myatt, G. J. (2007). *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Praveena, S., & Singh, S. P. (2014). Hybrid Clustering Algorithm and Feed-Forward Neural Network for Satellite Image Classification. In *2015 International Conference on Industrial Instrumentation and Control (ICIC)* (pp. 1378–1383). Pune: IEEE.
- Putri, S. A., & Wahono, R. S. (2015). Integrasi SMOTE dan Information Gain pada Naive Bayes untuk Prediksi Cacat Software. *Journal of Software Engineering*, 2(1).
- Radhimeenakshi, S. (2016). Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 3107–3111). New Delhi.
- Riquelme, J. C., Ruiz, R., Rodriguez, D., & Moreno, J. (2008). Finding defective modules from highly unbalanced datasets. In *Actas de los Talleres de las Jornadas de Ingeniería del Software y Bases de Datos* (Vol. 2, pp. 67–74). gijon, espana: Sistedes.
- Tiwari, D. (2014). Handling Class Imbalance Problem Using Feature Selection. *International Journal of Advanced Research in Computer Science & Technology*, 516–520.
- Unal, Y., & Kocer, E. (2013). Diagnosis of Pathology on the Vertebral Column with Backpropagation and Naive Bayes Classifier. *Technological Advances in Electrical, Electronics and Computer Engineering*, 278–281.
- WHO. (2007). *ICF-CY (International Classification of Functioning, Disability and Health for Children and Youth)*. Switzerland: World Health Organization.
- WHO. (2018). Disabilities. Retrieved May 4, 2018, from <http://www.who.int/news-room/facts-in-pictures/detail/disabilities>
- Yeh, Y., Hou, T., & Chang, W. (2012). Expert Systems with Applications An intelligent model for the classification of children ' s occupational therapy problems. *Expert Systems With Applications*, 39(5), 5233–5242. <https://doi.org/10.1016/j.eswa.2011.11.016>
- Zarchi, M. S., Bushehri, F., & Dehghanizadeh, M. (2018). SCADI: *International Journal of Medical Informatics*, 114(2), 1–28. <https://doi.org/10.1016/j.ijmedinf.2018.03.003>

- Zhang, D., Liu, W., Gong, X., & Jin, H. (2011). A Novel Improved SMOTE Resampling Algorithm Based on Fractal. *Journal of Computational Information Systems*, 6, 2204–2211.
- Zhang, Z.-Z., Chen, Q., Ke, S.-F., Wu, Y.-J., Qi, F., & Zhang, Y.-P. (2008). Ranking Potential Customers Based on Group-Ensemble. *International Journal of Data Warehousing and Mining*, 4(2), 79–89.
<https://doi.org/10.4018/jdwm.2008040109>