VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i1.4998

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

IMPROVING STUNTING CLASSIFICATION PERFORMANCE USING COMBINATION SMOTE TECHNIQUE AND ARTIFICIAL NEURAL NETWORK ALGORITHM

Wiga Maulana Baihaqi[1] ; Ida Nur Laela[2*]; Darso[3]

Technology Information[1,3]
System Information[2]
Universitas Amikom Purwokerto, Indonesia[1,2,3]
https://www.amikompurwokerto.ac.id/ [1,2,3]
wiga@amikompurwokerto.ac.id[1], laellaida719@gmail.com[2*], darso@amikompurwokerto.ac.id[3]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Child development is at the core of the nation's future. However, there are still serious problems that hinder optimal child growth, one of which is stunting. Stunting is a condition that has become a global concern in the context of public health and development. The use of deep learning algorithms has great potential to overcome the problem of stunting classification. The ratio of stunting handling is still a problem due to imbalance data. Classification algorithms such as ANN will experience a decrease in performance when faced with unbalanced classes, this makes it difficult to take action for early diagnosis of stunting. Synthetic Minority Oversampling Technique (SMOTE) is used to balance the failure data in diagnosis. The results showed that the combination of the SMOTE oversampling technique can improve the ability of the ANN algorithm model to accurately classify stunted or minority populations. The accuracy, precision, recall, and F1-Score values of this study are 0.90, 0.85, and 0.95, respectively. The results of (MCC) obtained a value of 0.73, and (G-Mean) of 0.86 before applying SMOTE and the results after applying SMOTE MCC of 0.84 and G-Mean of 0.92. This indicates that the minority group, namely stunted toddlers, can be predicted well. The implementation of the combination of SMOTE and ANN has proven successful in classifying imbalance stunting data, so this research can be used as a reference for future research to handle unbalanced data.*

Keywords: *artificial neural network, classification, deep learning, oversampling, stunting.*

Intisari - *Tumbuh kembang anak merupakan inti dari masa depan bangsa. Namun, masih ada masalah serius yang menghambat pertumbuhan anak secara optimal, salah satunya adalah stunting. Stunting merupakan kondisi yang telah menjadi perhatian global dalam konteks kesehatan dan pembangunan masyarakat. Penggunaan algoritma deep learning, memiliki potensi besar untuk mengatasi masalah klasifikasi stunting. Rasio penanganan stunting masih menjadi masalah karena data tidak seimbang (imbalance). Algoritma klasifikasi seperti ANN akan mengalami penurunan performa jika dihadapkan dengan kelas yang tidak seimbang, hal ini menghambat untuk melakukan tindakan diagnosis dini stunting. Synthetic Minority Oversampling Technique (SMOTE) digunakan untuk menyeimbangkan data kegagalan dalam diagnosis. Hasil penelitian menunjukkan bahwa kombinasi teknik Oversampling SMOTE dapat meningkatkan kemampuan model algoritma ANN dalam mengklasifikasikan populasi stunting atau minoritas secara akurat. Nilai akurasi, presisi, recall, dan F1-Score dari penelitian ini berturut-turut adalah 0,90, 0,85, dan 0,95. Hasil (MCC) diperoleh nilai sebesar 0.73, dan (G-Mean) sebesar 0.86 sebelum diterapkan SMOTE dan hasil setelah diterapkan SMOTE MCC sebesar 0.84 dan G-Mean sebesar 0.92. Hal ini mengindikasikan bahwa kelompok minoritas, yaitu balita yang mengalami stunting, dapat diprediksi dengan baik. Implementasi kombinasi SMOTE dan ANN terbukti berhasil melakukan klasifikasi pada data imbalance stunting, dengan begitu penelitian ini dapat dijadikan sebagai referensi bagi penelitian ke depan untuk menangani data tidak seimbang.*

Kata kunci: *jaringan syaraf tiruan, klasifikasi, pembelajaran mendalam, oversampling, pengerdilan.*

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.4998

INTRODUCTION

Children's growth and development is at the core of a nation's future. However, there are still serious problems that hinder the optimal growth of children, one of which is stunting. Stunting is a condition that has become a global concern in the context of public health and development. Stunting is a condition of stunted growth in the body of a child under five years old caused by chronic malnutrition[1][2]. Children who are stunted are shorter in height and do not reach the standard they should be[3]. This phenomenon is not only about physical growth delays, but also involves broader impacts on the well-being and productivity of individuals in later life. Stunting has the potential to slow down brain development, with long-term consequences such as mental retardation, lack of learning ability, and the risk of developing diabetes, hypertension, and obesity[4].

The World Health Organization (WHO) notes that stunting is one of the most important public health issues, especially in developing countries [5]. However, achieving significant long-term improvements in stunting requires cross-sectoral efforts and sustained cooperation. In the Indonesian Nutrition Status Survey (SSGI) at the BKKBN National Working Meeting, the prevalence of stunting in Indonesia was still 21.6%. This means that there are still toddlers who are stunted who need to be handled quickly for early detection of stunting.

To address stunting, a holistic and data-driven approach is needed. Information and communication technologies, such as the us of deep learning algorithms, have great potential to address various public health issues, including stunting in children under five. Deep learning techniques are used in AI to improve automated detection capabilities [6]. With the ability to process large and complex data, machine learning and deep learning algorithms can be used to create systems that can diagnose stunting automatically and effectively[7].

Artificial Neural Network (ANN) is one of the various Deep Learning algorithms. ANN is a computational model based on biological science that has processing components called neurons and connections/networks between them called coefficients or weights[8]. The reason for this is that the algorithm can learn data in a complex way. ANN consistently performs well when handling noisy or incomplete data. They can create models with complex relationships between variables and constants, ANNs can also build good results even with incomplete data[9]. Unbalanced data ratio is still a problem in classification algorithms difficult to diagnose. This problem greatly affects the predictive performance of classification algorithm models because models tend to predict with high accuracy for classes that have a larger number of classes.

Research on unbalanced classes has been widely carried out, including Fitriani et.al proposing the application of the naïve bayes algorithm with oversampling techniques [10], Pratama et.al proposed the application of data imbalance selection with RFECV and ADASYN Mutmainah and Siti propose handling unbalanced data in the case of possible stroke disease [11].

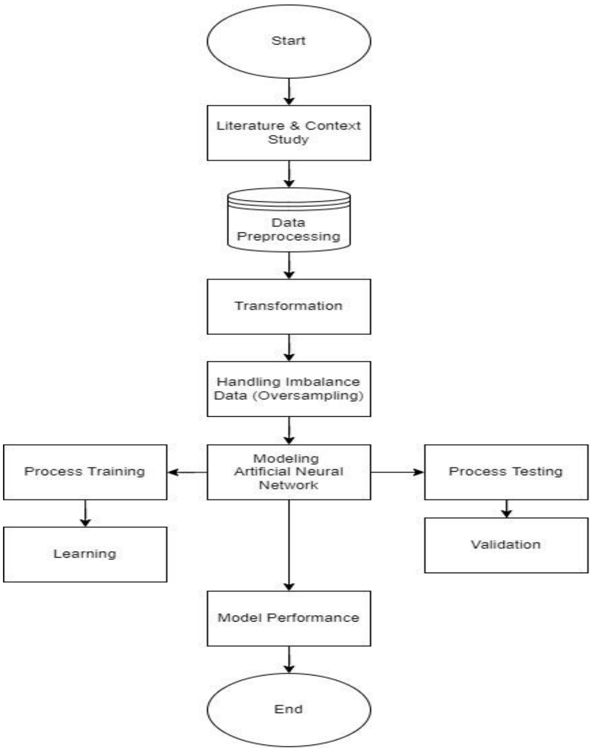
Data is said to be unbalanced if the proportion of minority classes is higher than that of other classes. The class that has more instances is referred to as the majority class, and the class with fewer instances is referred to as the minority class [12]. Some existing oversampling methods to handling unbalanced data are SMOTE, Random Sampling, Adaptive Synthetic Sampling (ADASYN), Random Over Sampling Examples (ROSE), etc. In this study, the SMOTE method was chosen because compared to the Random Sampling, ADASYN, and ROSE methods, SMOTE is simpler and easier to implement [13]. SMOTE is computationally much more efficient than the other methods. SMOTE also generates synthetic samples based on close neighbors in the feature space, while other methods only perform random doubling of minority class instances, which may increase the risk of overfitting[14]. Research on handling imbalance data in classification algorithms with SMOTE has been carried out which obtained an accuracy of NN with SMOTE 85%, DT with SMOTE 89%, LR with SMOTE 80%, and SVM 90% [15].

There are three models in the SMOTE technique that are used to address data imbalance undersampling, oversampling and combined method. Undersampling and oversampling data is used in the first method to balance the data distribution. The majority data is lowered in the undersampling technique to match the minority class, while the oversampling strategy creates new data for the minority class to match the majority class[7]. The second approach makes changes to the algorithm, such as increasing the minority class's weight. The third method combines methods that balance data distribution and algorithms [16].

Based on the background, the purpose of this study is to determine the success rate of the ANN algorithm and the SMOTE (Synthetic Minority Over-sampling Technique) oversampling technique in performing stunting diagnosis. The dataset used in this study is the result of the OTS stunting recap report in Patikraja District, Banyumas Regency, Central Java. This method was chosen because there is an unbalanced proportion in the dataset, where the number of examples belonging to different classes varies greatly.

MATERIALS AND METHODS

In this study to classify stunting using the artificial neural network algorithm, it can be seen through the research flow contained in the figure.



Source: (Research Results, 2024)
Figure 1. Research Diagram

The flow diagram below can be explained in several steps, which are as follows:

1. Literature & Context Study
   The first stage carried out to implement the research objectives is a literature study. At this stage, the process of formulating the problem and determining the strategy that will be used to conduct the analysis by conducting studies on previous research is carried out. This stage is carried out as a theoretical framework for decision support in utilizing datasets and using algorithms that will be used for stunting classification[17].
2. Data Processing
   This stage is the second stage of the research stage, namely data preprocessing using jupyter notebook and google colab tools. At this stage the data is processed to simplify the data starting from the data cleaning process or deleting data that has missing values or data that is not needed, data selection or data selection process.
3. Transformation
   This stage aims to change the type of data selected to facilitate the classification process with the artificial neural network algorithm. Changing the type of data used at this step will help the artificial neural network algorithm during the classification phase. There must be a transformation step since the data used in the classification process is numerical data. A label encoder is utilized to convert this data.
4. Handling Imbalance Data (Oversampling)
   Data imbalance handling is a set of techniques or strategies used to overcome the significant difference in sample size between the majority class and the minority class in a dataset[18]. Data imbalance can be a problem in the context of classification tasks, where one of the classes has significantly less representation than the other. The SMOTE method is one of the solutions in dealing with unbalanced data[19]. This method is carried out by increasing the number of minority classes so that they are equal to the majority classes by using raw data. Data is created or synthesized based on ANN models.
5. Modeling
   Modeling is a classification process that uses an artificial neural network algorithm with the Sequential method. The classification process consists of two processes the teaching process and the testing process. The model used is an artificial neural network training model for deep learning Artificial Neural Network (ANN) algorithm.
6. Model Performance
   At this point, the model is evaluated. The goal of evaluation is to match the model that is generated so that it matches the desired outcome. In this case, the developed model is an Artificial Neural Network algorithm. To understand the accuracy and performance threshold that are affected, this algorithm will be evaluated using the Confusion Matrix diagram[20].
   Where TP (True Positive) is the number of positive classes (stunting) correctly predicted as positive classes (stunting), FP (False Positive) is the number of negative classes (not stunting) that are incorrectly predicted as positive classes (stunting), TN (True Negative) is the real negative class (not stunting) and is correctly predicted as a negative class (not stunting), FN (False Negative) is the number of positive

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.4998

classes (stunting) that are incorrectly predicted as negative classes (stunting).
The confusion matrix calculation also produces accuracy, precision, recall, and f-1 score values.

The explanation and calculation are as follows:

1) Accuracy is something that describes how strong a model is in classifying data correctly.
   Accuracy = (TP + TN) / (TP_FP+FN+TN)
2) Precision is something that describes the accuracy between the data and the prediction results given by the model.
   Precision = (TP) / (TP+FP)
3) Recall is something that describes how well the model finds data again or the success of the model in finding back information.
   Recall = TP / (TP + FN)
4) F-1 Score describes the weighted average comparison of precision and recall.
   F-1 Score = (2* Recall * Precision) / (Recall + Precision)

RESULTS AND DISCUSSION

Data Preprocessing

The stunting dataset was collected from a dataset of the result of monitoring the nutritional status in the Patikraja sub-district of Banyumas district originating from the Puskesmas Patikraja. This dataset contains information on various characteristics of children and variables that may influence stunting.

Dataset has 348 variables consisting of children in the Patikraja sub-district, which consists of 13 villages. The data attributes in the dataset are gender, age, birth weight, birth length, body weight, body length, and stunting. The dataset was manually labeled for each stunting category based on the stunting prevalence index. From the recorded data, 59 children were identified as stunted and 289 children were identified as not stunted.

Table 1. Description Attributes Dataset Stunting

No	Attributes	Type	Description
1	Gender	Object	Indicates the gender of the child (for example, female, male).
2	Age	Int	Is the age of the child in weeks
3	Birth Weight	Object	Child's body weight at birth
4	Birth Length	Object	Child's body length at birth
5	Body Weight	Float	Child's weight at the time of data collection
6	Body Length	Float	Child's length at the time of data collection
7	Stunting	Int	This is a target variable or label that indicates whether the child is stunted or not.

Source: (Research Results, 2024)

In data preprocessing there is a process of cleaning data, checking missing values and checking correlations between variables and count data. There are no missing values in the data records. The following is checking the correlation between variables. View correlation between variables.

```
# Melihat korelasi antar variabel
corr = df.corr()
corr.style.background_gradient(cmap=
'coolwarm').set_precision(3)
```

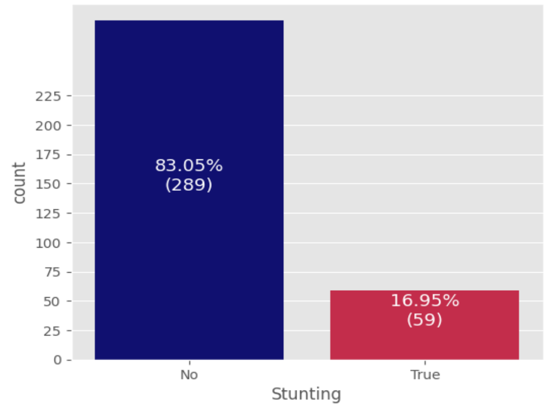
After checking, it can be seen in the following figure that there is a correlation between each variable.

Table 2. Correlation Variables

	Age	Body Weight	Body Length	Stunting
Age	1.000	0.787	0.902	-0.076
Body Weight	0.787	1.000	0.899	-0.306
Body Length	0.902	0.899	1.000	-0.267
Stunting	-0.076	-0.306	-0.267	1.000

Source: (Research Results, 2024)

Can be seen in table 2, That indicates a significant correlation within each variable. Following this, the researchers proceeded to examine the number of instances for both stunting and non-stunting data variables using Google Colab.



Source: (Research Results, 2024)
Figure 2. Show Diagram Stunting

The obtained diagram reveals an imbalanced distribution within the dataset. The majority class

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i1.4998

corresponds to "no stunting," while the "stunting" category represents the minority class.

Transformation
In this research, transformation is the process of normalizing data to a common size so that each feature has a similar probability of determining its classification results or estimating its dominance [15]. In transforming the data, the label encoder method is used which can be seen in below.

```
# Transformation category data to numerical
label_encoder = LabelEncoder()
df['Gender'] =
label_encoder.fit_transform(df['Gender'])
df['Stunting'] =
label_encoder.fit_transform(df['Stunting'])
```

Through the data transformation process, specific changes were introduced, particularly in the gender and stunting categories. The gender data was modified, assigning category 1 for males and the stunting category, while category 0 was assigned for females and the non-stunting category. This transformation effectively converted the original categorical data into numerical form, facilitating further analysis.

Table 3. Dataset Stunting

Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length	Stunting
0 1	37	2.3	38	11.5	90.0	1
1 0	56	3.6	49	13.6	95.5	0
2 1	30	3.2	46	10.4	84.0	0
3 1	52	3.1	50	14.5	96.5	0
4 0	52	2.8	49	12.0	93.0	0

Source: (Research Results, 2024)

Handling Imbalance Data
The technique used in handling unbalanced data is the SMOTE (Synthetic Minority Over-sampling Technique) there are Oversampling technique. This technique is performed before the data is tested in the modeling process with the ANN algorithm.

```
# SMOTE
# x is atrtibute influence stuntig
# y label is self
from imblearn.over_sampling import
SMOTE
```

```
x = df.drop(['Stunting'], axis=1)
y = df['Stunting']
```

Following the resampling process, the dataset was split into training and testing data. A noticeable distinction in the data volume is observed before and after resampling, as depicted in below.

```
X train shape:  (278, 6)
Y train shape:  (278,)
X test shape:  (70, 6)
Y test shape:  (70,)

X train shape:  (462, 6)
Y train shape:  (462,)
X test shape:  (116, 6)
Y test shape:  (116,)
```

Modeling
Data training with artificial neural network is done to obtain ANN model. The ANN algorithm used is a Sequential model. The architecture ANN used in this classification study is the first layer of 256 neurons, batch normalization layer to stabilize and speed up training and hidden layer, dropout layer with a dropout rate of 0.3 and one output layer for the stunting class.

```
# Create model ANN
model = keras.models.Sequential([
    keras.layers.Dense(256,
activation='relu',
input_dim=X_train.shape[1]),
    keras.layers.BatchNormalization(),
    keras.layers.Dropout(0.2),
    keras.layers.Dense(256,
activation='relu'),
    keras.layers.BatchNormalization(),
    keras.layers.Dropout(0.2),
    keras.layers.Dense(1,
activation='sigmoid'),
])

model.summary()
```

During training, the model undergoes a maximum of 100 iterations (epochs), with updates occurring every 16 samples based on the specified batch size. A validation split of 0.2 is implemented, meaning 20% of the training data serves as validation data to assess model performance during training. The highest accuracy for the test results is observed to be 0.9595 before oversampling and 0.9350 after oversampling, accompanied by a loss rate of 0.1127

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.4998

before oversampling and 0.1936 after oversampling. These results are achieved with a hidden layer comprising 100 nodes.

Model Performance

To comprehend the accuracy and performance thresholds affected, the algorithm will be assessed using the Confusion Matrix diagram. The figure below illustrates the model performance generated from the ANN model both before and after oversampling.

Table 4. Before Oversampling

	Precision	Recall	F1-Score	Support
0	0.89	0.98	0.93	57
1	0.86	0.46	0.60	13
acccuracy			0.89	70
macro avg	0.87	0.72	0.77	70
wighted avg	0.88	0.89	0.87	70

Source: (Research Results, 2024)

Table 5. After Oversampling

	Precision	Recall	F1-Score	Support
0	0.94	0.85	0.89	60
1	0.85	0.95	0.90	56
acccuracy			0.90	116
macro avg	0.90	0.90	0.90	116
wighted avg	0.90	0.90	0.90	116

Source: (Research Results, 2024)

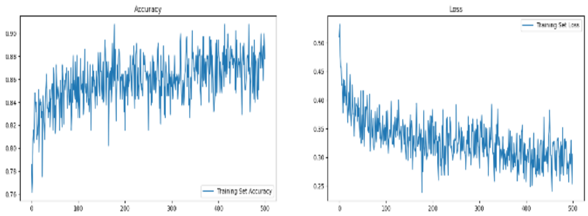
For ease of reading, the following table shows the differences between each model.

Table 6. Description of Model Results

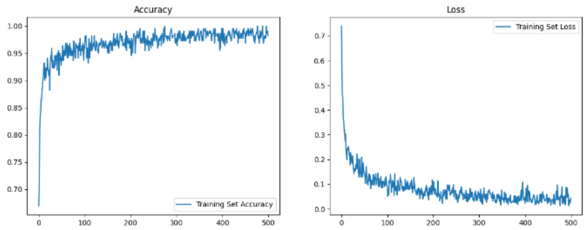
No		Model Artificial Neural Network	
1	Before Oversampling	Accuracy	0.89
		Precision	0.86
		Recall	0.46
		F-1 Score	0.60
2	After Oversampling	Accuracy	0.90
		Precision	0.85
		Recall	0.95
		F-1 Score	0.90

Source: (Research Results, 2024)

Artificial neural network before oversampling the ability to predict stunting is low, which has an accuracy value of 0.89, and has a precision, recall, and F-1 Score level of 0.86, 0.46 and 0.60 in predicting the stunting class. The recall and precision values obtained result in a low stunting class which means, that the model tends to miss stunting cases or (false negative) which can have a serious impact. While the results of the artificial neural network model after oversampling managed to increase accuracy, precision, recall and F1-Score. Where the accuracy value is 0.90, precision is 0.85, recall is 0.95 and F1-Score is 0.90. To see the level of accuracy and loss in the model can be seen in the following graph:



Source: (Research Results, 2024)
Figure 2. Before Oversampling



Source: (Research Results, 2024)
Figure 3. After Oversampling

It can be seen that the model does not experience overfitting after applying the SMOTE technique. From the training results before oversampling, the Matthews Correlation Coefficient (MCC) result is 0.73, and Geometric Mean (G-Mean) is 0.86 and the results after MCC is 0.84 and G-Mean is 0.92, to see the difference can be seen in the table below.

Table 7. MCC and G-Mean

No	Artificial Neural Network Model	MCC	G-Mean
1	Before	73%	86%
2	After	84%	92%

This shows the ability of the artificial neural network method to be successfully performed for stunting classification, and Oversampling successfully predicts the stunting class quite high.

CONCLUSION

This research uses the Oversampling SMOTE method to handle the imbalance. Based on the results of the above research, it is known that the use of the Oversampling SMOTE (Synthetic Minority Over-sampling Technique) method in the Artificial Neural Network model can avoid overfitting on unbalanced data and increase the classification accuracy for minority data. The data that is used is stunted data using an Artificial Neural Network algorithm. Artificial Neural Network can successfully classify stunting but cannot handle unbalanced data. Based on the analysis results from this study, the Oversampling SMOTE technique can improve the ability of the algorithm model to accurately classify stunted or minority populations. Data 348 record data were oversampled to create 578 data. This study's accuracy, precision, recall, and F1-Score values are 0.90, 0.85, and 0.95, respectively, indicating that minority groups, namely those experiencing stunting, may be predicted, the Matthews Correlation Coefficient (MCC) result is 0.73, and Geometric Mean (G-Mean) is 0.86 and the results after MCC is 0.84 and G-Mean is 0.92. This indicates that the oversampling technique is highly effective in achieving data imbalance correction. The use of Artificial Neural Network (ANN) algorithms and SMOTE in classifying stunting provides a reference for future research exploring the application of artificial intelligence in child health. This research can be used as a reference for future research to handle unbalanced data.

REFERENCE

[1] M. Trisiswati, D. Mardhiyah, dan S. Maulidya Sari, “Hubungan Riwayat Bblr (Berat Badan Lahir Rendah) Dengan Kejadian Stunting Di Kabupaten Pandeglang,” *Maj. Sainstekes*, vol. 8, no. 2, hal. 061–070, 2021, doi: 10.33476/ms.v8i2.2096.

[2] A. Rahmidini, “Hubungan stunting dengan perkembangan motorik dan kognitif anak,” *Semin. Nas. Kesehat.*, vol. 2, no. 1, hal. 90–104, 2020, [Daring]. Tersedia pada: http://www.ejurnal.stikesrespati-tsm.ac.id/index.php/semnas/article/download/272/192

[3] S. Syahrial, R. Ilham, Z. F. Asikin, dan S. S. I. Nurdin, “Stunting Classification in Children’s Measurement Data Using Machine Learning Models,” *J. La Multiapp*, vol. 3, no. 2, hal. 52–60, 2022, doi: 10.37899/journallamultiapp.v3i2.614.

[4] T. Vaivada, N. Akseer, S. Akseer, A. Somaskandan, M. Stefopulos, dan Z. A. Bhutta, “Stunting in childhood: An overview of global burden, trends, determinants, and drivers of decline,” *Am. J. Clin. Nutr.*, vol. 112, hal. 777S-791S, 2020, doi: 10.1093/ajcn/nqaa159.

[5] Kementrian Kesehatan RI, *PP Menteri Kesehatan RI*, vol. 21, no. 1. hal. 1–9. 2020,

[6] R. Resmiati dan T. Arifin, “Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination,” *Sistemasi*, vol. 10, no. 2, hal. 381, 2021, doi: 10.32520/stmsi.v10i2.1238.

[7] I. Ayuningtyas dan E. U. Kasanah, “Penerapan Synthetic Minority Oversampling Technique (Smote) Pada Kasus Dampak Covid-19 Terhadap (Synthetic Minority Oversampling Technique Approach in Case of the Impact,” *Bestari Bul. Stat. dan Apl. Terkini*, vol. I, hal. 1–7, 2021.

[8] G. R. Yang dan X. J. Wang, “Artificial Neural Networks for Neuroscientists: A Primer,” *Neuron*, vol. 107, no. 6, hal. 1048–1070, 2020, doi: 10.1016/j.neuron.2020.09.005.

[9] M. A. Rahman, R. chandren Muniyandi, D. Albashish, M. M. Rahman, dan O. L. Usman, “Artificial neural network with Taguchi method for robust classification model to improve classification accuracy of breast cancer,” *PeerJ Comput. Sci.*, vol. 7, hal. 2–27, 2021, doi: 10.7717/PEERJ-CS.344.

[10] R. D. Fitriani, H. Yasin, dan T. Tarno, “Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal),” *J. Gaussian*, vol. 10, no. 1, hal. 11–20, 2021, doi: 10.14710/j.gauss.v10i1.30243.

[11] S. Mutmainah, “Penanganan Imbalance Data Pada Klasifikasi Kemungkinan Penyakit Stroke,” *SNATi*, vol. 1, no. 1, hal. 10–16, 2021, doi: 10.20885/snati.v1i1.2.

[12] N. Javaid, N. Jan, dan M. U. Javed, “An adaptive synthesis to handle imbalanced big data with deep siamese network for electricity theft detection in smart grids,” *J. Parallel Distrib. Comput.*, vol. 153, hal. 44–52, 2021, doi: 10.1016/j.jpdc.2021.03.002.

[13] A. J. Mohammed, “Improving Classification Performance for a Novel Imbalanced Medical Dataset using SMOTE Method,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, hal. 3161–3172, 2020, doi: 10.30534/ijatcse/2020/104932020.

[14] D. Dablain, B. Krawczyk, dan N. V. Chawla, “DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 9, hal. 6390–6404, 2023, doi: 10.1109/TNNLS.2021.3136503.

[15] N. Chamidah, M. Mega Santoni, dan N. Matondang, “Terakreditasi SINTA Peringkat 2 Pengaruh Oversampling pada Klasifikasi Hipertensi dengan Algoritma Naïve Bayes, Decision Tree, dan Artificial Neural Network (ANN),” *Masa Berlaku Mulai*, vol. 1, no. 3, hal. 635–641, 2021.

[16] K. M. Hasib *dkk.*, “A Survey of Methods for Managing the Classification and Solution of Data Imbalance Problem,” *J. Comput. Sci.*, vol.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.4998

16, no. 11, hal. 1546–1557, 2020, doi: 10.3844/JCSSP.2020.1546.1557.

[17] M. Y. Matdoan, U. A. Matdoan, dan M. Saleh Far-Far, "Algoritma K-Means Untuk Klasifikasi Provinsi di Indonesia Berdasarkan Paket Pelayanan Stunting," *PANRITA J. Sci. Technol. Arts*, vol. 1, no. 2, hal. 41–46, 2022, [Daring]. Tersedia pada: https://journal.dedikasi.org/pjsta

[18] P. Vuttipittayamongkol, E. Elyan, dan A. Petrovski, "On the class overlap problem in imbalanced data classification," *Knowledge-Based Syst.*, vol. 212, no. 106631, 2021, doi: 10.1016/j.knosys.2020.106631.

[19] B. Kovács, F. Tinya, C. Németh, dan P. Ódor, "Unfolding the effects of different forestry treatments on microclimate in oak forests: results of a 4-yr experiment," *Ecol. Appl.*, vol. 30, no. 2, hal. 321–357, 2020, doi: 10.1002/eap.2043.

[20] D. Krstinić, M. Braović, L. Šerić, dan D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," hal. 01–14, 2020, doi: 10.5121/csit.2020.100801.