Combining Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI) and Hybrid Sampling in Handling Multi-Class Imbalance and Overlapping

Hartono
Universitas IBBI

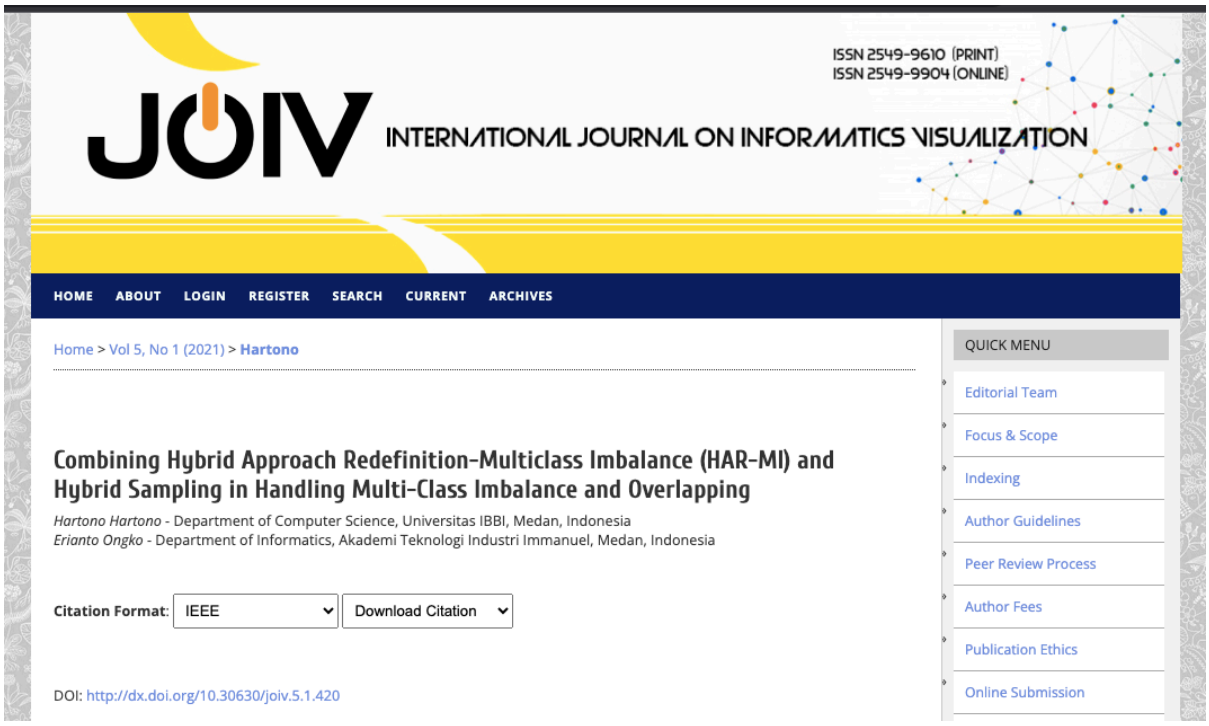
Erianto Ongko
Akademi Teknologi Industri Immanuel

JOIV : International Journal on Informatics Visualization Vol. 5 No. 1, Maret 2021

1. Submit pada OJS Jurnal : 4 November 2020
2. Accepted and Review Result: 29 November 2020
3. Pembayaran dan Submit hasil revisi : 30 November 2020
4. Pelaksanaan ICAITI 2020: 4 Desember 2020
5. Copyediting dan terbit: 20 Maret 2021

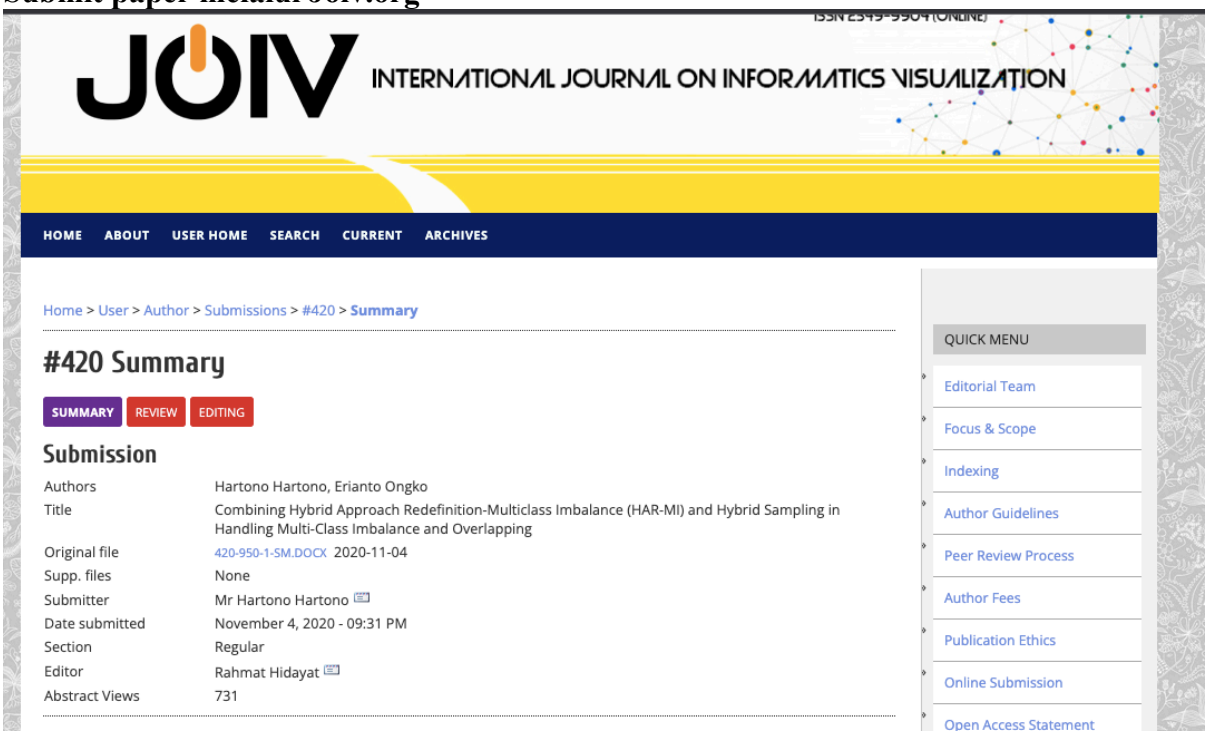
Bukti Korespondensi Pengajuan Guru Besar Hartono

Judul Karya Ilmiah	:	Combining Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI) and Hybrid Sampling in Handling Multi-Class Imbalance and Overlapping
Jurnal	:	JOIV : International Journal on Informatics Visualization
Penulis	:	Hartono, Erianto Ongko
Edisi	:	2021
Volume	:	Vol. 5 No. 1
Penerbit	:	Politeknik Negeri Padang



Bukti Bahwa Keikutsertaan pada ICAITI 2020 dimana semua paper yang diterima dan dipresentasikan akan dipublikasikan pada jurnal JOIV

Submit paper melalui Joiv.org



Gmail

Hartono Ibbi <hartonoibbi@gmail.com>

Notification Accepted for 3rd ICAITI

4 pesan

International Journal On Informatics Visualization <joiv@pnp.ac.id> 29 November 2020 pukul 10.16
Kepada: hartonoibbi@gmail.com, eriantoongko@gmail.com

Dear Mr Hartono Hartono:

Congratulations - your paper #420 ('Combining Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI) and Hybrid Sampling in Handling Multi-Class Imbalance and Overlapping') for 3rd ICAITI 2020 has been accepted and will be presented at the ICAITI conference.

Please complete the registration and make the payment at this link http://bit.ly/icaitireg2020.

The payment must be made by December 2nd, 2020.
for more information about registration and payment, please follow this link http://icaiti.org/registration/

The conference will be held online on December 4th, 2020, using a Zoom meeting. We will inform all authors about the link before the day of the conference.

For publication in the International Journal on Informatics Visualization, Please submit your revision before January 10th, 2020. Please add the proofreader's certificate/letter on the last page that ensures your paper has been proofread.

The comment from a reviewer :

The general comment from this manuscript is needed to follow the format JOIV template which can be downloaded on the website. The following comments should be revised accordingly :

1. The number of words in the abstract should be a minimum of 220 words and a maximum of 250 words.

2. The Manuscript consists of 4 chapters : INTRODUCTION, MATERIAL AND METHOD, RESULT AND DISCUSSION, CONCLUSION.

3. The manuscript doesn't present well, please correct the layout, make them display nicely, balance them in two columns. Make sure the manuscript displays a fix in every page. Suggestion : save as in word (DOCX).

4. Please update your abstract into 220-250 words and your reference 70% in (2017-2020) from a journal indexed by Scopus. Citation and Reference in Paper must use Mendeley with IEEE Style.

5. The topic is suitable with our conference theme, but still needs some improvement in both language and format. Some writing that doesn't need to be underlined. Improve conclusions by adding quantitative results.

6. The manuscript proposed a method handling the imbalance dataset and overlapping. The author proposed the HAR-Mi method to solve the problem. The manuscript is suitable to present in icaiti 2020.

International Journal on Informatics Visualization
http://joiv.org/index.php/joiv

Hartono Ibbi <hartonoibbi@gmail.com> 30 November 2020 pukul 10.00
Kepada: International Journal On Informatics Visualization <joiv@pnp.ac.id>

Dear ICAITI 2020 Committee,

We have made a registration payment and have filled in the Google Form at http://bit.ly/icaitireg2020. We also attach proof of payment in this email. Please provide information on whether the proofreader's certificate / letter can be filled in by the Head of the English Language Center at our University. Thank you for the opportunity for us to participate in ICAITI 2020 and if there is anything we need to complete, please do not hesitate to inform us.

Sincerely Yours,

Hartono
[Kutipan teks disembunyikan]

Transfer ICAITI Hartono.PNG
360K

International Journal On Informatics Visualization <joiv@pnp.ac.id> 30 November 2020 pukul 10.21
Kepada: Hartono Ibbi <hartonoibbi@gmail.com>

Dear Author

Hartono Ibbi

Thank you for joining ICAITI 2020

We have received your registration.

We have no specific format about proofreader certificates.
Every provider has its own format. It generally consists of manuscript title, author, date, name of provider.
It's just to show the manuscript was edited with proper english.
you can attach it to the last page of your revisions. for more detail about the publication process in JOIV journal we will explain in conference after parallel session.

Best Regard

JOIV

[Kutipan teks disembunyikan]

Hartono Ibbi <hartonoibbi@gmail.com> 30 November 2020 pukul 14.53
Kepada: International Journal On Informatics Visualization <joiv@pnp.ac.id>

Dear ICAITI 2020 Committee,

Thank you for your fast response and well noted with thanks.

Best regards,

Hartono
[Kutipan teks disembunyikan]

PERBAIKAN YANG DILAKUKAN SESUAI KOMENTAR DARI REVIEWER

Combining Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI) and Hybrid Sampling in Handling Multi-Class Imbalance and Overlapping

Hartono[1,2], Erianto Ongko[3*]

[1] *Department of Computer Science, Universitas IBBI, Medan, 20114, Indonesia*
[2] *Department of Computer Science, Universitas Potensi Utama, Medan, 20241, Indonesia*
E-mail: hartonoibbi@gmail.com

[3]*Department of Informatics, Akademi Teknologi Industri Immanuel, 20114, Medan, Indonesia*
E-mail: eriantoongko@gmail.com

Abstract—**The class imbalance problem in the multi-class dataset is more difficult to handle than the problem in the two class and this problem is more complicated if accompanied by overlapping. One method that has proven reliable in dealing with this problem is the Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI) method which is classified as a hybrid approach that combines sampling and classifier ensembles. However, in terms of diversity among classifiers, hybrid approach that combines sampling and classifier ensembles will give better results. HAR-MI provides excellent results in handling multi-class imbalances. The HAR-MI method uses SMOTE to increase the number of sample in minority class. However, this SMOTE also has a weakness where if there is an extremely imbalanced dataset and a large number of attributes there will be over-fitting. In order to overcome the problem of over-fitting, the Hybrid Sampling method was proposed. HAR-MI combination with Hybrid Sampling is done to increase the number of samples in the minority class and at the same time reduce the number of noise samples in the majority class. The preprocessing stages at HAR-MI will use the Minimizing Overlapping Selection under Hybrid Sampling (MOSHS) method and the processing stages will use Different Contribution Sampling. The results obtained will be compared with the results using Neighbourhood-based undersampling. Overlapping and Classifier Performance will be measured using Augmented R-Value, the Matthews Correlation Coefficient (MCC), Precision, Recall, and F-Value. The results showed that HAR-MI with Hybrid Sampling gave better results in terms of Augmented R-Value, Precision, Recall, and F-Value.**

Keywords— **Class Imbalance; Multi-Class Dataset; Multi-Class Imbalance; Hybrid Approach; HAR-MI.**

Comment [Office1]: The number of words in the abstract should be a minimum of 220 words and a maximum of 250 words.

I. INTRODUCTION

The problem of class imbalance has become one of the most interesting problems in data mining[1]. Class imbalance has become one of the most interesting issues to research regarding data mining, machine learning, and knowledge discovery[2]. This problem occurs because most of the real-world dataset is in an imbalance state and if it is not handled properly it will cause a class with a small number of samples to become unrepresented and reduce the level of accuracy[3]. In general, the approach to solving class imbalance problems can be divided into 3 (three), namely: data-level, algorithm-level, and hybrid[4]. The data-level approach focuses on efforts to change the distribution of data through a process of over sampling or under sampling. Over sampling was carried out on the minority class and under sampling was carried out on the majority class[5]. On the other hand, the algorithm-level approach does not change the distribution of data, but focuses on classifier efforts to pay more attention to minority classes through the application of bagging, boosting, or through the ensemble process of existing classifiers[6].

Hybrid Approach is an approach that combines Data-Level and Algorithm-Level[7]. In terms of diversity and classifier performance, a hybrid approach that combines the use of sampling and classifier ensembles will give good results[8]. In fact, the Hybrid Method is not only good at dealing with binary-class imbalance, but also multi-class imbalance problems[9]. Multi-class imbalance problems are more difficult to handle than binary-class imbalance and usually multi-class balance problems do not stand alone, but are accompanied by overlapping[10]. This problem becomes even more challenging if the minority classes are in overlapping conditions[11].

Comment [Office2]: The Manuscript consists of 4 chapters : INTRODUCTION, MATERIAL AND METHOD, RESULT AND DISCUSSION, CONCLUSION.

To minimize the impact of multi-class imbalance which is accompanied by overlapping, the preprocessing process has a very significant effect[12]. For this problem, feature selection method is often used at the preprocessing stage, so the effort to apply the preprocessing stage in the hybrid approach is a wise choice[13]. One of the methods in the hybrid approach that applies preprocessing and gives satisfactory results in this problem is the Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI)[14].

As with most methods in the hybrid approach, HAR-MI also uses the oversampling method for minority classes by using SMOTE in the feature selection process at the preprocessing stage. One of the Feature Selection methods that provides excellent results in handling overlapping is Minimizing Overlapping Selection under SMOTE (MOSS)[15], even though this oversampling process often causes overfitting[16]. In addition, other problems that are often found in the application of SMOTE are overgeneralization and noise[17]. The use of Minority Over-Sampling Techniques (M-SMOTE) and Edited Nearest Neighbor (ENN), which are a type of Hybrid Sampling, has yielded very satisfying results[18].

It would be interesting if there is a method that combines multi-class balance handling followed by overlapping and at the same time paying attention so that the sampling process does not over fit. This study will combine the use of HAR-MI with Hybrid Sampling. The results of this study will be compared with Neighbourhood-based undersampling which is one of the best methods of handling multi-class imbalance and overlapping[19].

II. The Material and Method

A. Hybrid Approach

The pseudocode of the Hybrid Approach is as follows[20].

$Input: D_T = \{x_1, x_2, \dots, x_n\}$ //Training Dataset
$N = Number\ of\ Classifier$
$Output: Classification\ Prediction\ P$
$Method:$
$Step\ 1\ Preprocessing\ using\ Preprocessing\ Method$
$Step\ 2\ For\ i = 1\ to\ N\ do$
  $i. Apply\ Machine\ Learning\ Classification\ Algorithm\ on\ The\ Attributes\ of\ D_T$
  $ii. Obtain\ Classification\ Prediction\ P_i\ from\ machine\ learning\ classification\ algorithm$
$End\ For$
$Step\ 3\ For\ i = 1\ to\ n$
  $Apply\ processing\ using\ bagging, boosting\ or\ sampling$
$End\ For$

B. Hybrid Sampling

The pseudocode of the Hybrid Sampling using M-SMOTE and ENN is as follows[18].

$Input: Dataset\ S, Minority\ Samples\ S_{Min}, Majority\ Sample\ S_{Maj}$
$Output: Final\ Dataset\ S'$
$Create\ global\ variable\ G_{max}, create\ array\ Eva_{min}, Eva_{maj}, Eva$
$Step\ 1: If\ G_{MCC} = 0$
  $Processing\ M - SMOTE\ for\ S_{Min}$
  $Processing\ ENN\ for\ S_{Maj}$
  $End\ If$
$Step\ 2: Calculate\ Eva_{min}\ using\ MCC$
  $Calculate\ Eva_{maj}\ using\ MCC$
  $Calculate\ Eva\ using\ MCC$
$Step\ 3: If Eva_{min} < Eva\ or\ If Eva_{maj} < Eva$
  $G_{MCC} = G_{MCC} - 1$
  $End\ If$
$Step\ 4: if\ G_{MCC} < 0$
  $Terminate\ and\ Output\ Final\ Dataset\ S'$
  $else$
  $Return\ to\ Step\ 1$
  $End\ If$

C. Augmented R-Value

Augmented R-Value states how much overlapping occurs. The greater the Augmented R-Value, the greater the overlapping[21].

$$R_{aug}(D[V]) = \frac{\sum_{i=0}^{k-1} |C_{k-1-i}| R(C_i)}{\sum_{i=0}^{k-1} |C_i|} \quad (1)$$

Where $C_0, C_1, \dots, C_{k-1}$ are k class labels with $|C_0| \geq |C_1| \geq \cdots \geq |C_{k-1}|$ and $D[V]$: Dataset D containing predictors in set V. Larger R_{Aug} is higher overlap degree of a dataset.

D. Classifier Performance

Classifier Performance will be measured using the Matthews Correlation Coefficient (MCC), Precision, Recall, and F-Value. This classifier performance measurement is carried out based on the confusion matrix which can be seen in Table 1[22].

TABLE I
Confusion Matrix

	Predictive Positive Class	**Predictive Negative Class**
Actual Positive Class	True Positive (TP)	False Negative (FN)
Actual Negative Class	False Positive (FP)	True Negative (TN)

The Matthews Correlation Coefficient (MCC), Precision, Recall, and F-Value calculations can be seen in the following equation[18].

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TN \times FN)(TN \times FP)(TN \times FN)(TP \times FP)}} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = TP \quad (4)$$

$$F - Value = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

E. Proposed Method / Algorithm

The research stages can be seen in Figure 1.

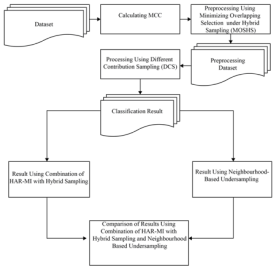


Fig. 1 Research Stage

Comment [Office3]: The Manuscript consists of 4 chapters : INTRODUCTION, MATERIAL AND METHOD, RESULT AND DISCUSSION, CONCLUSION.

F. *Preprocessing Using Minimizing Overlapping Selection under Hybrid Sampling (MOSHS)*

The pseudocode of the preprocessing stage is as follows.

$X - matrix\ with\ p\ predictors: X = [x_1, x_2, \dots, x_p]; class\ label: y$
For All Samples in Minority
    Hybrid Sampling the Minority Class using $m - SMOTE$
End For
Create NewMinority
For All Samples in Majority
    Hybrid Sampling the Majority Class using ENN
End For
Create NewMajority
ForAll Samples in NewMinority and NewMajority
    Preprocessed Dataset
End For

G. *Processing Using Different Contribution Sampling* (DCS)

The pseudocode of the processing stage is as follows.

For $i = 1$ *to Number of Instance in Preprocessed Dataset*
  Add Preprocessed Dataset to S_i
  $B - SVM$ *will do for Classifying* S_i
  Determine the Majority Class
  Determine the Minority Class
  For All Instance in Majority Class
    NewSVSets[ ] will form by checking and delete the noise in SVSets
    NewNSVSets[ ] will form by Multiple Hybrid Sampling
  End For
  For All Instance from NewSVSets and NSVSets
    Create an instance for majority class
  End For
  For All Instance in minority class
    SMOTEBoost Process for SVSets and Create SMOTESets
  End For
  For All SMOTESets and NewNSVSets do
    NewPositiveSampleSets
  End For
  For All NewNegativeSampleSets and NewPositiveSampleSets do
    ResultDataSet
  End For
End For

III. Results and Discussion

A. *Dataset Description*

The multi-class imbalanced datasets used in this study were sourced from the KEEL Repository[23]. The dataset used can be seen in Table II.

TABLE II
DATASET DESCRIPTION

Dataset	#Ex	#Atts	Distribution of Class	IR
Contraceptive	1473	9	629/333/511	1.89
Flare	1066	11	147/211/239/95/43 /331	7.70
Car Evaluation	1728	6	384/69/1210/65	18.62
Thyroid Disease	720	21	17/37/666	39.18
Red Wine Quality	1599	11	10/53/681/638/199 /18	68.10
Page-Blocks	5473	10	4913/329/28/88/ 115	188.72

In Table II, it can be seen that the dataset used has various imbalance ratios, ranging from low, medium, and high imbalance ratio. Likewise, the number of samples also varied.

B. *Testing Result*

The first test was conducted to obtain Augmented R-Value and MCC values. The test results can be seen in Table III.

TABLE III
TESTING FOR AUGMENTED R-VALUE AND MCC

Dataset	HAR-MI with Hybrid Sampling		Neighbourhood Based Undersampling	
	Augmented R-Value	MCC	Augmented R-Value	MCC
Contraceptive	0.327	0.97	0.337	0.91
Flare	0.357	0.83	0.359	0.82
Car Evaluation	0.367	0.85	0.373	0.81
Thyroid Disease	0.379	0.81	0.381	0.79
Red Wine Quality	0.411	0.75	0.415	0.71
Page-Blocks	0.436	0.73	0.437	0.71

Based on Table III, it can be seen that for the Augmented R-Value the results obtained by HAR-MI with Hybrid Sampling are better than the Neighbourhood-based undersampling. The greater the Augmented R-Value, the greater the overlapping that occurs. Based on the Augmented R-Value obtained by the two methods, it can be seen that the greater the imbalance ratio value, the greater the tendency for overlapping to occur. The MCC value provided by HAR-MI With Hybrid Sampling is also better than that obtained by Neighbourhood-based undersampling.

The second test was conducted to obtain Precision, Recall, and F-Value. The test results can be seen in Table IV.

TABLE IV
TESTING FOR PRECISION, RECALL, AND F-VALUE

Dataset	HAR-MI with Hybrid Sampling			Neighbourhood Based Undersampling		
	Precision	Recall	F-Value	Precision	Recall	F-Value
Contraceptive	0.88	0.97	0.92	0.78	0.89	0.83
Flare	0.85	0.88	0.87	0.81	0.87	0.84
Car Evaluation	0.84	0.89	0.86	0.76	0.73	0.75
Thyroid Disease	0.87	0.76	0.81	0.85	0.71	0.77
Red Wine Quality	0.82	0.81	0.81	0.82	0.72	0.77
Page-Blocks	0.78	0.77	0.77	0.77	0.69	0.73

Based on Table IV, it can be seen that based on the Precision, Recall, and F-Value values the results given by HAR-MI with Hybrid Sampling are better than the results obtained by Neighbourhood-based undersampling.

Comment [Office4]: The Manuscript consists of 4 chapters : INTRODUCTION, MATERIAL AND METHOD, RESULT AND DISCUSSION, CONCLUSION.

C. *Statistical Tests*

To validate the results of the study, a statistical test was conducted to measure performance using the Wilcoxon Signed-Rank Test[24]. The statistical test results can be seen in Table V.

TABLE V
STATISTICAL TESTS USING WILCOXON SIGNED-RANK TEST

Performance Measurement	P-Value	Hypothesis
Augmented R-Value	0.0355223	*H_0 (no significant score difference between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling) rejected and this means H_1 (there is a significant difference between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling in score) Accepted because the p-value <0.05*
MCC	0.0355223	*H_0 (no significant score difference between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling) rejected and this means H_1 (there is a significant difference between HAR-MI with Hybrd Sampling and Neighbourhood-Based Undersampling in score) Accepted because the p-value <0.05*
Precision	0.0625000	*H_0 (no significant score difference between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling) is accepted and this means H_1 (there is a significant difference between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling in score) is rejected because the p-value >0.05*
Recall	0.0312500	*H_0 (no significant score difference between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling) rejected and this means H_1 (there is a significant difference between HAR-MI with Hybrd Sampling and Neighbourhood-Based Undersampling in score) Accepted because the p-value <0.05*
F-Value	0.0340064	*H_0 (no significant score difference between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling) rejected and this means H_1 (there is a significant difference between HAR-MI with Hybrd Sampling and Neighbourhood-Based Undersampling in score) Accepted because the p-value <0.05*

D. *Discussion*

Based on the test results and Statistical Tests, it can be seen that in terms of overlapping the HAR-MI method with Hybrid Sampling gives better results compared to MCC between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling. However, in general the results obtained in overlapping handling are good, where the Augmented R-Value obtained is not too high. Augmented R-Value is very dependent on the imbalance ratio, the higher the value of the imbalance ratio, the higher the overlapping that occurs. Based on statistical tests, it can be seen that there is a significant difference for Augmented R-Value and MCC between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling.

As for the MCC value, the results given by HAR-MI with Hybrid Sampling are still better and there is a tendency that the more classes there are, the lower the MCC value obtained. As for the Precision, Recall, and F-Value values, the results obtained show that HAR-MI with Hybrid Sampling is also better than MCC between HAR-MI with Hybrid Sampling and Neighbourhood-Based Undersampling. The results obtained indicate that the higher the imbalance ratio, the value of Precision, Recall, and F-Value obtained also decreases.

Based on the results of statistical testing with the Wilcoxon Signed-Rank Test, it was found that for Augmented R-Value, the P-Value is 0.0355223, the P-Value for MCC is 0.0355223, the P-Value for Recall is 0.0312500, and the P-Value for F -Value is 0.0340064. This means that for Augmented R-Value, MCC, Recall, and F-Value there is a significant difference between the results given by HAR-MI with Hybrid Sampling and Neighborhood-Based Undersampling. As for Precision, although the results given by HAR-MI are better than Neighborhood-Based Undersampling, but based on the test results with the Wilcoxon Signed-Rank Test there is no significant difference as indicated by the P-Value obtained> 0.05, where the P-Value obtained is 0.0625000.

IV. CONCLUSIONS

Based on the results in Tables III, IV, and V, it can be seen that in terms of handling multi-class imbalance and overlapping, the results obtained by using HAR-MI with Hybrid Sampling give better results when compared to

Comment [Office5]: The Manuscript consists of 4 chapters : INTRODUCTION, MATERIAL AND METHOD, RESULT AND DISCUSSION, CONCLUSION.

Neighbourhood-Based Undersampling. The results obtained show that HAR-MI with Hybrid Sampling excels at all test values such as Augmented R-Value, MCC, Precision, Recall, and F-Value.

This shows that for handling multi-class imbalance, Hybrid Sampling, which can avoid over fitting, also gives better results compared to Under Sampling or Over Sampling. Future Research, can pay attention to the handling of multi-class imbalance accompanied by overlapping in a state of high yield ratio and also on datasets with a large number of classes and a large number of attributes.

ACKNOWLEDGMENT

This research was funded by a Grant from the Ministry of Education and Culture and the Ministry of Research and Technology of the Republic of Indonesia.

REFERENCES

[1] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from Class-Imbalanced Data: Review of Methods and Applications," *Expert Systems With Applications*, vol. 73, pp. 220–239, May 2017.

[2] A. Guzmán-Ponce, J. S. Sánchez, R. M. Valdovinos, and J. R. Marcial-Romero, "DBIG-US: A two-stage under-sampling algorithm to face the class imbalance problem," *Expert Systems with Applications*, p. 114301, Nov. 2020, doi: 10.1016/j.eswa.2020.114301.

[3] B. Liu and G. Tsoumakas, "Dealing with class imbalance in classifier chains via random undersampling," *Knowledge-Based Systems*, vol. 192, p. 105292, Mar. 2020, doi: 10.1016/j.knosys.2019.105292.

[4] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J Big Data*, vol. 6, no. 1, p. 27, Mar. 2019, doi: 10.1186/s40537-019-0192-5.

[5] P. Shamsolmoali, M. Zareapoor, L. Shen, A. H. Sadka, and J. Yang, "Imbalanced data learning by minority class augmentation using capsule adversarial networks," *Neurocomputing*, Jul. 2020, doi: 10.1016/j.neucom.2020.01.119.

[6] W. Hou, X. Wang, H. Zhang, J. Wang, and L. Li, "A novel dynamic ensemble selection classifier for an imbalanced data set: An application for credit risk assessment," *Knowledge-Based Systems*, vol. 208, p. 106462, Nov. 2020, doi: 10.1016/j.knosys.2020.106462.

[7] F. Rayhan, S. Ahmed, A. Mahbub, R. Jani, S. Shatabda, and D. M. Farid, "CUSBoost: Cluster-Based Under-Sampling with Boosting for Imbalanced Classification," in *2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, Dec. 2017, pp. 1–5, doi: 10.1109/CSITSS.2017.8447534.

[8] J. Zhao, J. Jin, S. Chen, R. Zhang, B. Yu, and Q. Liu, "A weighted hybrid ensemble method for classifying imbalanced data," *Knowledge-Based Systems*, vol. 203, p. 106087, Sep. 2020, doi: 10.1016/j.knosys.2020.106087.

[9] Z. Liu, D. Tang, Y. Cai, R. Wang, and F. Chen, "A hybrid method based on ensemble WELM for handling multi class imbalance in cancer microarray data," *Neurocomputing*, vol. 266, pp. 641–650, Nov. 2017, doi: 10.1016/j.neucom.2017.05.066.

[10] E. R. Q. Fernandes and A. C. P. L. F. de Carvalho, "Evolutionary inversion of class distribution in overlapping areas for multi-class imbalanced learning," *Information Sciences*, vol. 494, pp. 141–154, Aug. 2019, doi: 10.1016/j.ins.2019.04.052.

[11] Y. Zhu, Y. Yan, Y. Zhang, and Y. Zhang, "EHSO: Evolutionary Hybrid Sampling in overlapping scenarios for imbalanced learning," *Neurocomputing*, vol. 417, pp. 333–346, Dec. 2020, doi: 10.1016/j.neucom.2020.08.060.

[12] P. Zyblewski, R. Sabourin, and M. Woźniak, "Preprocessed dynamic classifier ensemble selection for highly imbalanced drifted data streams," *Information Fusion*, vol. 66, pp. 138–154, Feb. 2021, doi: 10.1016/j.inffus.2020.09.004.

[13] L. Yijing, G. Haixiang, L. Xiao, L. Yanan, and L. Jinling, "Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data," *Knowledge-Based Systems*, vol. 94, pp. 88–104, Feb. 2016, doi: 10.1016/j.knosys.2015.11.013.

[14] H. Hartono, Y. Risyani, E. Ongko, and D. Abdullah, "HAR-MI method for multi-class imbalanced datasets," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, Art. no. 2, Apr. 2020, doi: 10.12928/telkomnika.v18i2.14818.

[15] G.-H. Fu, Y.-J. Wu, M.-J. Zong, and L.-Z. Yi, "Feature selection and classification by minimizing overlap degree for class-imbalanced data in metabolomics," *Chemometrics and Intelligent Laboratory Systems*, vol. 196, p. 103906, Jan. 2020, doi: 10.1016/j.chemolab.2019.103906.

[16] X. Gao *et al.*, "An ensemble imbalanced classification method based on model dynamic selection driven by data partition hybrid sampling," *Expert Systems with Applications*, vol. 160, p. 113660, Dec. 2020, doi: 10.1016/j.eswa.2020.113660.

[17] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, and D. Huang, "New imbalanced bearing fault diagnosis method based on Sample-characteristic Oversampling TechniquE (SCOTE) and multi-class LS-SVM," *Applied Soft Computing*, vol. 101, p. 107043, Mar. 2021, doi: 10.1016/j.asoc.2020.107043.

[18] Z. Xu, D. Shen, T. Nie, and Y. Kou, "A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data," *Journal of Biomedical Informatics*, vol. 107, p. 103465, Jul. 2020, doi: 10.1016/j.jbi.2020.103465.

[19] P. Vuttipittayamongkol and E. Elyan, "Neighbourhood-based undersampling approach for handling imbalanced and overlapped data," *Information Sciences*, vol. 509, pp. 47–70, Jan. 2020, doi: 10.1016/j.ins.2019.08.062.

[20] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, Jul. 2012, doi: 10.1109/TSMCC.2011.2161285.

[21] S. Oh, "A new dataset evaluation method based on category overlap," *Comput. Biol. Med.*, vol. 41, no. 2, pp. 115–122, Feb. 2011, doi: 10.1016/j.compbiomed.2010.12.006.

[22] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, Jul. 2019, doi: 10.1016/j.patcog.2019.02.023.

[23] J. Alcalá-Fdez *et al.*, "KEEL: a software tool to assess evolutionary algorithms for data mining problems," *Soft Comput*, vol. 13, no. 3, pp. 307–318, Feb. 2009, doi: 10.1007/s00500-008-0323-y.

[24] F. Wilcoxon, "Individual Comparisons by Ranking Methods on JSTOR," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

Comment [Office6]: Improve conclusions by adding quantitative results.

Comment [Office7]: your reference 70% in (2017-2020) from a journal indexed by Scopus. Citation and Reference in Paper must use Mendeley with IEEE Style.

CERTIFICATE OF PROOFREADING

No. 009/SPn/R/IBBI/I/2021

This document certified that the paper listed below was reviewed dan edited for proper English language, grammar, punctuation, spelling, and overall style by the undersigned. Neither the research content nor the authors intention were altered in any way during the editing process.

PAPER TITLE

"Combining Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI) and Hybrid Sampling in Handling Multi-Class Imbalance and Overlapping"

AUTHORS

HARTONO
ERIANTO ONGKO

DATE ISSUED

January 7, 2021

Dr. John Pieter Simarmata, S.S., M.Hum.
Head of English Laboratory of Universitas IBBI

Registration and Payment Confirmation Form ICAITI 2020
1 pesan

Google Forms <forms-receipts-noreply@google.com> 30 November 2020 pukul 09.53
Kepada: hartonoibbi@gmail.com

Thanks for filling out Registration and Payment Confirmation Form ICAITI 2020

Here's what we got from you:

Registration and Payment Confirmation Form ICAITI 2020

Bank Information
Bank : Bank Negara Indonesia (BNI)
Bank Branch : Padang
Bank Account : 390804108
Account Name : RPL 010 PTN Padang utk PS KJS

Email address *

hartonoibbi@gmail.com

Full Name *

Hartono

Phone Number *

085270601999

Type of Registration *

- ( ) International Presenter
- (•) Domestic Presenter

Paper id *

If you are a presenter, please enter your paper title

#420

Title of Paper *

If you are a presenter, please enter your paper title

Combining Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI) and Hybrid Sampling in Handling Multi-Class Imbalance and Overlapping

Other People Registered in the Payment *

If you are registering and paying for more than one person all at once, please mention the person's name. Please use new line for each person.

-

Payment Confirmation *

- (•) Bank Transfer

Name of Account

Hartono

Name of Bank

Danamon

Bank Account Number

98337959

Amount of Payment

4000000

Date of Payment Made

MM DD YYYY

11 / 30 / 2020

Proof of payment

Submitted files

Transfer ICAITI Hartono - Hartono Ibbi.PNG

3rd International Conference on Applied Information Technology and Innovation (3rd ICAITI 2020)

Schedule overview

Friday, 04 December 2020

Main Room (exclude parallel session):

https://bit.ly/icaiti-mainroom
Meeting ID: 951 8242 0289
Passcode: icaiti2020

07.30 – 08.15 WIB	**Registration**	**Committee**
08.15 – 08.25	Welcome and Introduction	**MC**
08.25 – 08.35	Welcome Remark By **Alde Alanda** **Chair of 3rd ICAITI**	**MC**
08.35 – 08.50	Welcome Remark By **Dr. Surfa Yondri, ST, S.ST, M.Kom** **Director of Politeknik Negeri Padang**	**MC**
08.50 – 09.00	**Break/preparation for keynote speaker**	**Committee**
9.00 – 9.45	**Keynote Speaker** **Prof. S.G Ahmed** Zagazig University, Egypt Formal representative as Associate Editor North Africa and Middle East, AHRO Publisher	Moderator: **Fazrol Rozi**
9.45-10.30	**Keynote Speaker** **Assoc. Prof. Ts Dr Hairul Nizam Mahdin** Universiti Tun Hussein Onn, Malaysia	Moderator: **Deddy Prayama**
10.30-12.00	**Parallel Session** 1. 3rd ICAITI A 2. 3rd ICAITI B 3. 3rd ICAITI C 4. 3rd ICAITI D 5. 3rd ICAITI E	Session Chair : **1. Deddy Prayama** **2. Hidra Amnur** **3. Aldo Erianda** **4. Alde Alanda** **5. Fazrol Rozi**
	Break	
14.30-15.15	**Keynote Speaker**	Moderator:

	Prof. Dr-Ing Hendro Wicaksono Jacobs University Bremen, Germany	**Dr. Deden Witarsyah**
15.15-16.00	**Keynote Speaker** **Assoc. Prof. Leonardo Alfonso** IHE Delft Institute for Water Education, Netherlands	Moderator: **Dr. Nurhamidah**
	Break	
16.30-17.00	**Greeting from JOIV and explanation for article publication guideline** **by Mr. Rahmat Hidayat** Editor in Chief of JOIV	**MC**
17.00-17.15	**4th ICAITI 2021 Logo Launching** **By Mr. Rahmat Hidayat** Editor in Chief of JOIV	**Committee**
17.15-17.30	**Closing Remark** **By Ronal Hadi** Head of Information Technology Department Politeknik Negeri Padang	**MC**
17.30-	**Photo Session**	**Committee**
	Finish	

Parallel Session 3rd ICAITI A

Session Chair: Deddy Prayama

Zoom link:

https://bit.ly/icaiti-room1

Meeting ID: 964 9930 2390

Passcode: icaiti2020

No	ID Paper	Title	Author
1	431	Single Image Dehazing Using Deep Learning	Hartanto, Rahadianti
2	448	Role Comparison between Deep Belief Neural Network and NeuroEvolution of Augmenting Topology to Detect Diabetes	Wijaya
3	449	THE COMPREHENSIVE OF MAMDANI INFERENCE TO SUPPORT THE DECISION OF SCHOLARSHIP GRANTEE CASE	Rahmayuni
4	460	Design and Implementation of a User-Centered Web-App using Open Source Platform: Indonesia Disaster Data (InDITA)	Rudiastuti, Suryanegara, Wirawan,...
5	461	Drivers of Cloud Computing Adoption in Small Medium Enterprises of Indonesia Creative Industry	Gui, Fernando, Mokhtar, Shaharudin,...
6	476	A Systematic Literature Review of Different Machine Learning Methods on Hate Speech Detection	Rudy Salim, Suhartono
7	420	Combining Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI) and Hybrid Sampling in Handling Multi-Class Imbalance and Overlapping	Hartono, Ongko

JOIV INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION
ISSN 2549-9610 (PRINT)
ISSN 2549-9904 (ONLINE)

#420 Editing

SUMMARY REVIEW **EDITING**

Submission

Authors	Hartono Hartono, Erianto Ongko
Title	Combining Hybrid Approach Redefinition-Multiclass Imbalance (HAR-MI) and Hybrid Sampling in Handling Multi-Class Imbalance and Overlapping
Section	Regular
Editor	Rahmat Hidayat

Copyediting

COPYEDIT INSTRUCTIONS

REVIEW METADATA	REQUEST	UNDERWAY	COMPLETE
1. Initial Copyedit File: None	—	—	—
2. Author Copyedit File: None Choose File No file chosen Upload	—	—	
3. Final Copyedit File: None	—	—	—

Copyedit Comments No Comments

Layout

Galley Format	FILE		
1. PDF VIEW PROOF	420-1241-1-PB.PDF	2021-03-20	307

Supplementary Files	FILE
	None

Layout Comments No Comments

Proofreading

REVIEW METADATA

	REQUEST	UNDERWAY	COMPLETE
1. Author	—	—	
2. Proofreader	—	—	—
3. Layout Editor	—	—	—

Proofreading Corrections No Comments PROOFING INSTRUCTIONS

JOIV : International Journal on Informatics Visualization
ISSN **2549-9610 (print) | 2549-9904 (online)**
Organized by Society of Visual Informatics, and Institute of Visual Informatics - UKM and Soft Computing and Data Mining Centre - UTHM
W : http://joiv.org
E : joiv@pnp.ac.id, hidra@pnp.ac.id, rahmat@pnp.ac.id

View JOIV Stats

» Submission Received: **March 3, 2020**
» Submission Accepted: **July 30, 2020**
» **SCOPUS CiteScore Tracker 2020**

» **WoS / Web of Science**
» Latest submission: September 16, 2018
» **Web of Science** Citation Analysis

» **IET INSPEC**
» Added to review: **May 29, 2020**

» **Ei COMPENDEX**
» Submission: **February 10, 2021**

PUBLICATION PARTNERS

USER

You are logged in as...
hartono
» My Profile
» Log Out

AUTHOR

Submissions
» Active (0)
» Archive (2)
» New Submission