Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem

Hartono
Universitas Potensi Utama

Erianto Ongko
Akademi Teknologi Industri Immanuel

Dahlan Abdullah
Universitas Malikussaleh

International Journal of Advances in Intelligent Informatics Vol. 7 No. 3, November 2021

1. Paper has been registered : 29 Juni 2019
2. Hasil Review Accepted with Major Revision: 5 Agustus 2019
3. Upload hasil revisi : 11 Agustus 2019
4. Accepted with Minor Revision: 9 November 2021
5. Upload hasil revisi: 22 November 2021
6. Accepted: 22 November 2021

Bukti Korespondensi Pengajuan Guru Besar Hartono

Judul Karya Ilmiah	:	Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem
Jurnal	:	International Journal of Advances in Intelligent Informatics
Penulis	:	Hartono, Erianto Ongko, Dahlan Abdullah
Edisi	:	2021
Volume	:	Vol. 7 No. 3
Penerbit	:	Universitas Ahmad Dahlan

Bukti Submit Pertama pada International Symposium on Advanced Intelligent Informatics (SAIN) dimana All Accepted and Presented Paper akan dipublikasikan pada IJAIN

Gmail

Hartono Ibbi <hartonoibbi@gmail.com>

[SAIN'19] Paper 1570568696 has been uploaded

1 pesan

sain@ijain.org <sain@ijain.org> 29 Juni 2019 pukul 12.55
Cc: Andri Pranolo <andri.pranolo@tif.uad.ac.id>, Hartono Hartono <hartonoibbi@gmail.com>, Erianto Ongko <eriantoongko@gmail.com>, Dahlan Adbullah <dahlan.unimal@gmail.com>

Dear Mr. Hartono Hartono:

Thank you for uploading your paper 1570568696 ('Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem') to 2019 2nd International Symposium on Advanced Intelligent Informatics (SAIN). The paper is of type application/pdf and has a length of 912979 bytes.

You can modify your paper at https://edas.info/showPaper.php?m=1570568696 and see all your submissions at https://edas.info/r26041 using the EDAS identifier hartonoibbi@gmail.com

Regards,
The Technical program chairs

Gmail

Hartono Ibbi <hartonoibbi@gmail.com>

[SAIN'19] Paper 1570568696 has been registered

1 pesan

sain@ijain.org <sain@ijain.org> 29 Juni 2019 pukul 10.35
Kepada: Hartono Hartono <hartonoibbi@gmail.com>
Cc: Andri Pranolo <andri.pranolo@tif.uad.ac.id>

Dear Mr. Hartono Hartono:

Thank you for registering your paper 1570568696 ('Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem') to 2019 2nd International Symposium on Advanced Intelligent Informatics (SAIN). You still have to upload your manuscript at https://edas.info/uploadPaper.php?m=1570568696. Your manuscript can be application/pdf and application/msword.

You can see all your submissions and their status at

https://edas.info/index.php?c=26041

using your EDAS user id hartonoibbi@gmail.com.

Once you upload your manuscript, you will receive another email confirmation.

Regards,
The Technical program chairs

HASIL REVIEW ACCEPTED

Gmail

Hartono Ibbi <hartonoibbi@gmail.com>

[SAIN'19] Your paper #1570568696 ('Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem')

1 pesan

sain@ijain.org <sain@ijain.org> 5 Agustus 2019 pukul 04.57
Balas Ke: sain@ijain.org
Kepada: Hartono Hartono <hartonoibbi@gmail.com>, Erianto Ongko <eriantoongko@gmail.com>, Dahlan Adbullah <dahlan.unimal@gmail.com>

Dear Mr. Hartono Hartono:

Congratulations! We are pleased to inform you that your paper #1570568696 ('Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem') has been accepted, but requires major revisions and will be presented in the SAIN'19.

The reviews are below or can be found at https://edas.info/showPaper.php?m=1570568696.

Your accepted and registered paper will be published in the Indexed Journal which will be informed when you have revised the article according to this notification and reviewer comments. Please noted that your paper might be rejected by the journal editors if the paper does not follow the reviewer comments and journal standard.

In order to guarantee your paper to be included in the journal, please do the following steps:

1. Revise your paper taking all the suggestions and comments of the reviewers into account.

2. The length of the submitted paper is at least 11 pages and no more than 15 pages in A4 Paper size and follow all requirements at http://sain.ijain.org/submission/

3. Articles in the section structure: Introduction - Related Works - Method - Results and Discussion – Conclusion

4. The paper must be in MS Word file format only.

5. Expect a minimum of 25 references primarily with a minimum of 60% to journal papers.

6. Please ensure that the similarity rating of your final manuscript must not exceed 10%.

7. Please complete your registration and payment. You may refer to http://sain.ijain.org/registration/. Please noted that your Early Bird Registration is August 12, 2019, and the Final registration is August 15, 2019. It can be made by PayPal ijain@uad.ac.id or Bank Transfer:

Bank Name : BANK RAKYAT INDONESIA
Account No : 0029 01 091282 50 8
Swift Code : BRINIDJA
Branch Address : BRI Katamso
City : Yogyakarta
Country : Indonesia
Account holder name : Andri Pranolo
ID/Passport Number : 3471030503830004/C0921966
Address : Gendeng GK IV/953 Yogyakarta 55225 Indonesia
Email : andri.pranolo@tif.uad.ac.id
Phone (Mobile) number: +6281392554050
----
8. Send your revision no later than August 12, 2019, through your EDAS account, for further evaluation to be fully accepted in the journal.

9. Please be informed that the journal editor has the right to contact the author directly should there be any further amendments for journal publication purpose.

Best Regards

The Technical program chairs

======= SAIN review_form_1 1 =======

> *** Relevance to the conference: - Is the paper closely related to the theme of the conference (broadly conceived)? - Is the content interesting enough to a broad audience? - Is the paper readable in a multi-disciplinary context?
Poor (1)

> *** Structure of the paper: Structure of the paper
Poor (1)

> *** Standard of English: Standard of English
Poor (1)

> *** Appropriateness of abstract as a description of the paper: Appropriateness of abstract as a description of the paper
Poor (1)

> *** Use and number of keywords/key phrases: Use and number of keywords/key phrases
Poor (1)

> *** Relevance and clarity of drawings, graphs and tables: Relevance and clarity of drawings, graphs and tables
Poor (1)

> *** Appropriateness of the research/study method: Appropriateness of the research/study method
Poor (1)

> *** Discussion and conclusions: Discussion and conclusions
Poor (1)

> *** Reference list, adequate and correctly cited: Reference list, adequate and correctly cited
Poor (1)

> *** Significance: - How important is the work reported? Does it attack an important/difficult problem (as opposed to a peripheral/simple one)? - Does the approach offered advance the state of the art? - Does it involve or synthesize ideas, methods, approaches from multiple disciplines? - Does it have interesting implications for multiple disciplines?*
Poor (1)

> *** Originality: - Is this a new issue? Is this a novel approach to an issue? - Is this a novel combination of familiar ideas/techniques/methods/approaches? - Does the paper point out differences from related research? - Does the paper properly situate itself with respect to previous work?
Poor (1)

> *** Quality: - Is the paper technically sound? How are its claims backed up? - Does it carefully evaluate the strengths and limitations of its contribution?
Poor (1)

> *** Clarity: - Is the paper clearly written? Does it motivate the research? Does it describe clearly the methods employed (e.g., experimental procedures, algorithms, analytical tools), if any? - Are the results, if any, described and evaluated thoroughly? - Is the paper organized in a sensible and logical fashion?
Poor (1)

> *** Comments: Explanations for the above ratings and other general comments to the authors

The paper has been published in (International Journal of Advances in Intelligent Informatics ISSN 2442-6571 Vol. 4, No. 2, July 2018, pp. xx-xx 1).

======= SAIN review_form_1 2 =======

> *** Relevance to the conference: - Is the paper closely related to the theme of the conference (broadly conceived)? - Is the content interesting enough to a broad audience? - Is the paper readable in a multi-disciplinary context?
Good (3)

> *** Structure of the paper: Structure of the paper
Good (3)

> *** Standard of English: Standard of English
Good (3)

> *** Appropriateness of abstract as a description of the paper: Appropriateness of abstract as a description of the paper
Good (3)

> *** Use and number of keywords/key phrases: Use and number of keywords/key phrases
Good (3)

> *** Relevance and clarity of drawings, graphs and tables: Relevance and clarity of drawings, graphs and tables
Good (3)

> *** Appropriateness of the research/study method: Appropriateness of the research/study method
Good (3)

> *** Discussion and conclusions: Discussion and conclusions
Poor (1)

> *** Reference list, adequate and correctly cited: Reference list, adequate and correctly cited
Good (3)

> *** Significance: - How important is the work reported? Does it attack an important/difficult problem (as opposed to a peripheral/simple one)? - Does the approach offered advance the state of the art? - Does it involve or synthesize ideas, methods, approaches from multiple disciplines? - Does it have interesting implications for multiple disciplines?*
Good (3)

> *** Originality: - Is this a new issue? Is this a novel approach to an issue? - Is this a novel combination of familiar ideas/techniques/methods/approaches? - Does the paper point out differences from related research? - Does the paper properly situate itself with respect to previous work?
Good (3)

> *** Quality: - Is the paper technically sound? How are its claims backed up? - Does it carefully evaluate the strengths and limitations of its contribution?
Good (3)

> *** Clarity: - Is the paper clearly written? Does it motivate the research? Does it describe clearly the methods employed (e.g., experimental procedures, algorithms, analytical tools), if any? - Are the results, if any, described and evaluated thoroughly? - Is the paper organized in a sensible and logical fashion?
Good (3)

> *** Comments: Explanations for the above ratings and other general comments to the authors

The research is very interesting, the paper is well structured and organized. Only two points to correct:

It is necessary to correct the following phrase: "majority class based on class lable"

Where are the conclusions and future work? Please include.

======= SAIN review_form_1 3 =======

> *** Relevance to the conference: - Is the paper closely related to the theme of the conference (broadly conceived)? - Is the content interesting enough to a broad audience? - Is the paper readable in a multi-disciplinary context?
Good (3)

> *** Structure of the paper: Structure of the paper
Good (3)

> *** Standard of English: Standard of English
Good (3)

> *** Appropriateness of abstract as a description of the paper: Appropriateness of abstract as a description of the paper
Good (3)

> *** Use and number of keywords/key phrases: Use and number of keywords/key phrases
Good (3)

> *** Relevance and clarity of drawings, graphs and tables: Relevance and clarity of drawings, graphs and tables
Good (3)

> *** Appropriateness of the research/study method: Appropriateness of the research/study method
Good (3)

> *** Discussion and conclusions: Discussion and conclusions
Good (3)

> *** Reference list, adequate and correctly cited: Reference list, adequate and correctly cited
Good (3)

> *** Significance: - How important is the work reported? Does it attack an important/difficult problem (as opposed to a peripheral/simple one)? - Does the approach offered advance the state of the art? - Does it involve or synthesize ideas, methods, approaches from multiple disciplines? - Does it have interesting implications for multiple disciplines?*
Good (3)

> *** Originality: - Is this a new issue? Is this a novel approach to an issue? - Is this a novel combination of familiar ideas/techniques/methods/approaches? - Does the paper point out differences from related research? - Does the paper properly situate itself with respect to previous work?
Good (3)

> *** Quality: - Is the paper technically sound? How are its claims backed up? - Does it carefully evaluate the strengths and limitations of its contribution?
Good (3)

> *** Clarity: - Is the paper clearly written? Does it motivate the research? Does it describe clearly the methods employed (e.g., experimental procedures, algorithms, analytical tools), if any? - Are the results, if any, described and evaluated thoroughly? - Is the paper organized in a sensible and logical fashion?
Good (3)

> *** Comments: Explanations for the above ratings and other general comments to the authors

It would be better if some more comprehensive evaluations are included.
The future work in this field to be mention.
No empirical relations used to validate techniques mention.

Authors Must Check before Submitting Final Paper:
Figures & tables should be cited in the text & the insertion of the same should be very near to where it is cited. Figures should be visible.
The research paper should be written in the perspective of the third person. Words such as ‘I’ ‘we’ ‘our’ etc. needs to be avoided.
Check the spelling mistakes & grammatical errors.

REVISI SESUAI MASUKAN DARI REVIEWER

Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem

Hartono [a,b,1], Erianto Ongko [c,2*], Dahlan Abdullah [d,3]

[a] Department of Computer Science, STMIK IBBI, Medan, Indonesia
[b] Department of Computer Science, Universitas Potensi Utama, Medan, Indonesia
[c] Department of Informatics, Akademi Teknologi Industri Immanuel, Medan, Indonesia
[d] Department of Informatics, Universitas Malikussaleh, Lhokseumawe, Indonesia
[1] hartonoibbi@gmail.com; [2] eriantoongko@gmail.com*; [3] dahlan@unimal.ac.id
* corresponding author

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Class Imbalance
Hybrid Approach Redefinition
Hybrid Ensembles
Classifier
Data Diversity

ABSTRACT

Class Imbalance problems often occur in the classification process, the existence of these problems is characterized by the tendency of a class to have instances that are much larger than other classes. This problem certainly causes a tendency towards low accuracy in minority classes with smaller number of instances and also causes important information on minority classes not to be obtained. Various methods have been applied to overcome the problem of the imbalance class. One of them is the Hybrid Approach Redefinition method which is one of the Hybrid Ensembles methods. The tendency to pay attention to the performance classifier, has led to an understanding of the importance of selecting an instance that will be used as a classifier. In the classic Hybrid Approach Redefinition method classifier selection is done randomly using the Random Under Sampling approach, and it is interesting to study how performance is obtained if the sampling process is based on Cluster-Based by selecting existing instances. The purpose of this study is to apply the Hybrid Approach Redefinition method with Cluster-Based Instance Selection (CBIS) approach so that it can obtain a better performance classifier. The results showed that Hybrid Approach Redefinition with cluster-based instance selection gave better results on the number of classifiers, data diversity, and performance classifiers compared to classic Hybrid Approach Redefinition.

1. Introduction

Class Imbalance problems are characterized by the presence of a class with a number of instances that are much smaller (minority class) and other classes with a much larger number of instances (majority class). This problem is a major problem in the classification process and has attracted the attention of researchers in the fields of data mining and machine learning[1]. If we discuss the problem of an imbalance class, the main consequence is low accuracy in minority classes, namely a class with a number of instances that are much smaller than majority classes, which is a class with a much larger number of instances[2]. Besides that, it should be noted that the classification process is carried out assuming that the distribution of instances in each class is the same, so if there is a problem of the imbalance class it will result in important information in a minority class with a much smaller number of instances that cannot be obtained[3].

Research on class imbalance problems has always been an interesting topic, especially if it is related to the problem of classification and machine learning which is very interesting to the attention of many researchers at this time. One method that draws the attention of researchers in overcoming the problem of this imbalance class is the hybrid approach[4]. Another approach that is widely used is data driven and algorithm driven[5]. The data driven and algorithm driven approach to handling imbalance classes experiences the main issue of losing important information and training data overfit. As for the hybrid ensembles the main issue is training time[6]. To reduce training time, the hybrid ensembles method in principle adopts the sampling principle which is combined with boosting[7][8]. However, there are other aspects that need to be considered in handling the imbalance class, which is related to the number of classifier and data diversity[9]. Another Hybrid Ensembles method, Hybrid Approach Redefinition, is one of the Hybrid Ensembles approaches based on sampling and boosting[10]. This method has been tested quite well in handling imbalance classes with a good number of classifiers and data diversity. However, what needs to be considered is the performance classifier, especially when faced with a dataset with a large number of attributes[11].

Basically the classification is based on the selection and placement of existing instances based on a number of existing classifiers[12]. This situation is the main thing that needs to be considered when dealing with the imbalance class problem in a dataset with a large number of attributes. Therefore, developing a Cluster-Based Instance Selection (CBIS) approach which is an Under Sampling method which is stated to be able to help well in the sampling process when there is a dataset with a large number of attributes[13]. The characteristics of clustering analysis with instance selection will combine and complement each other in the Under Sampling process for majority classes. The performance classifier commonly used in research on the imbalance class is the measurement of Sensitivity, Specificity, F-Measure, and G-Mean[14]. Based on a number of previous studies, this study will discuss the application of the Hybrid Approach Redefinition method with the Cluster-Based Instance Selection (CBIS) approach in handling imbalance class problems so that a better performance classifier can be obtained, especially when compared to Hybrid Approach Redefinition classic.

The rest of the paper is structured as follows. Section II presents the research method. Section III provides experimental process using Hybrid Approach Redefinition with Cluster-Based Instance Selection and Hybrid Approach Redefinition Classic. The experimental process and dataset used are presented in Section IV with the results. Finally, Section V concludes the paper and gives recommendations for future works.

2. Related Works

Sampling Method is one of the approaches in handling class imbalance. This process is done by generating a new dataset from a dataset that has a imbalance class where the new dataset has a better distribution balance between majority and minority classes[15]. In general, the sampling method can be divided into three groups: Under Sampling, Over Sampling, and Hybrid Methods. The Under Sampling method focuses on reducing samples from Majority Class, while the Over Sampling method focuses on adding samples from Minority Class[6]. In Under Sampling, the instance selection process will get better results compared to ones trained using the original dataset[16]. In connection with the selection of samples in Under Sampling, it is known that cluster-based sample selection will get better results than random sample selection[17]. Another problem arises when handling two-class imbalance, with Under Sampling which has a primary focus on the Majority Class, causing the instance selection process to experience constraints because basically the selection instance is designed to distinguish groups of samples in multiclass datasets, it is difficult to apply to samples that only exist in one class, namely majority class[18]. Cluster-Based Instance Selection (CBIS) approach is a combination of the Under Sampling method with the Instance Selection where the process instance selection will increase the ability of Under Sampling to Majority Class[13].

However, if handling imbalance classes only focus on the Majority Class, it will result in poor data diversity. Whereas handling class imbalance is expected to not only use a small number of classifiers but also obtain good data diversity[9]. Hybrid Approach Redefinition (HAR) Method offers handling class imbalance by using a small number of classifiers and good data diversity because the focus of handling is not only on the majority class but also on the minority class[10]. The Cluster-Based Instance Selection method is very interesting to be integrated in the Hybrid Approach Redefinition (HAR) Method. Especially in the process of Different Contribution

Sampling using the Biased Support Vector Machine in the Majority Class[7]. Hybrid Approach Redefinition (HAR) Method with Cluster-Based Instance Selection is expected to obtain a smaller number of classifier and better data diversity compared to the Classic Hybrid Approach Redefinition (HAR) Method.

3. Method

The study was conducted to test the number of classifiers, data diversity, and performance classifier. Performance classifiers are measured based on sensitivity, specificity, F-Measure, and G-Means. In this study a comparison between Hybrid Approach Redefinition and Cluster-Based Instance Selection (CBIS) will be carried out with classic Hybrid Approach Redefinition. The process will begin with preprocessing stages, processing stages, and evaluation stages. The experimental process in this study will be carried out using datasets sourced from the KEEL Dataset Repository using datasets that have a large number of attributes[19].

3.1. Preprocessing Stage

This preprocessing stage will carry out the process of selecting instances that will be used as classifiers using the Cluster-Based Instance Selection (CBIS) method and the SMOTEBoost Method. There is a slight difference with Hybrid Approach Redefinition classic which uses Random Under Sampling and SMOTEBoost methods. The preprocessing stages in the Hybrid Approach Redefinition with CBIS can be seen in Figure 1.

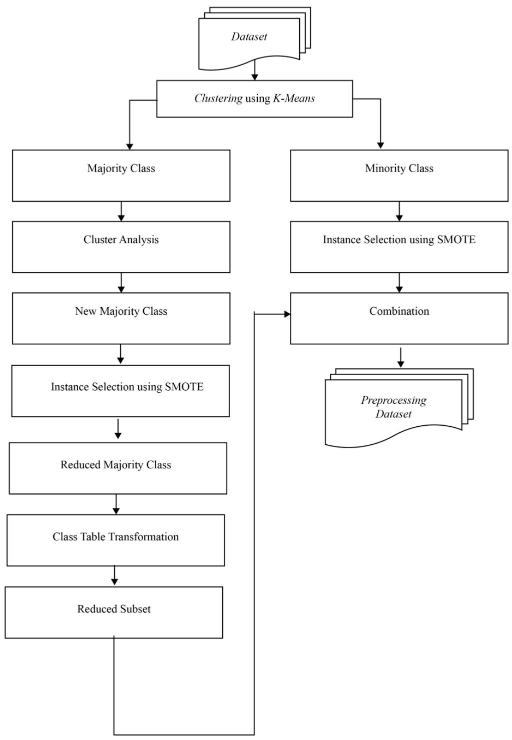

Fig. 1. Preprocessing Stage

In Figure 1 it can be seen that, if the clustering results indicate the existence of an imbalance class problem, which is characterized by the existence of certain classes with a large number of instances (majority class) and the presence of classes with a small number of instances (minority class). In majority classes a clustering analysis and instance selection process will be carried out. Clustering analysis will be conducted to group instances in majority classes where each instance belongs to a specific cluster, which is called a subclass of majority class. Then, every instance that exists will associate with a new class label, so that a new majority class will be generated. The next stage will be the process of instance selection using SMOTEBoost, where this process is intended to measure the size of the classifier to produce reduced majority class. Then the results will be transformed and recombined into reduced subset of majority ckass based on class lable information. While the Minority class will undergo a cluster analysis stage for grouping instances that exist in minority classes, and then will undergo the process of instance selection and the results will be combined with reduced subset of majority class to become a preprocessing dataset.

3.2. Processing Stage

The processing stage will be carried out using Biased Support Vector Machine. The Biased Support Vector Machine process will produce Support Vector Sets (SV Sets) and Non Support Vector Sets (NSV Sets) for both majority class and minority class. For NSV Sets in Majority Class under sampling process will be carried out using Cluster-Based Instance Selection while for SV Sets in Minority Class, instance selection will be processed using SMOTEBoost. The processing stages can be seen in Figure 2.

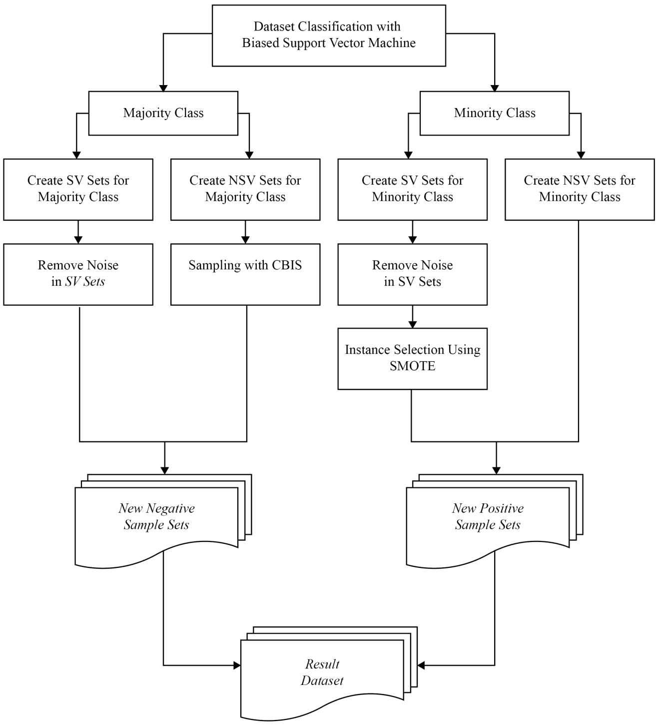

Fig. 2. Processing Stage

In Figure 2 it can be seen that preprocessing datasets originating from preprocessing stages will be classified using Biased Support Vector Machine to SV Sets and NSV Sets for Majority Class and Minority Class. In the next process the noise in the SV Sets Majority Class will be removed and then combined with the NSV Sets Majority Class which has undergone a sampling process using CBIS to become New Negative Sample Sets. Whereas NSV Sets in the minority class will be combined together with SV Sets in the majority class whose number has been removed and has undergone an instance selection process by using SMOTEBoost to become New Positive Sample Sets. Both New Negative Sample Sets and New Positive Sample Sets will be the Result dataset.

3.3. Evaluation Stage

The evaluation stage is intended to compare the results obtained between Hybrid Approach Redefinition with Cluster-Based Instance Selection (CBIS) and classic Hybrid Approach Redefinition. Evaluation is done by looking at a number of parameters such as the number of classifiers, diversity data, and performance classifier. The evaluation stages can be seen in Figure 3.

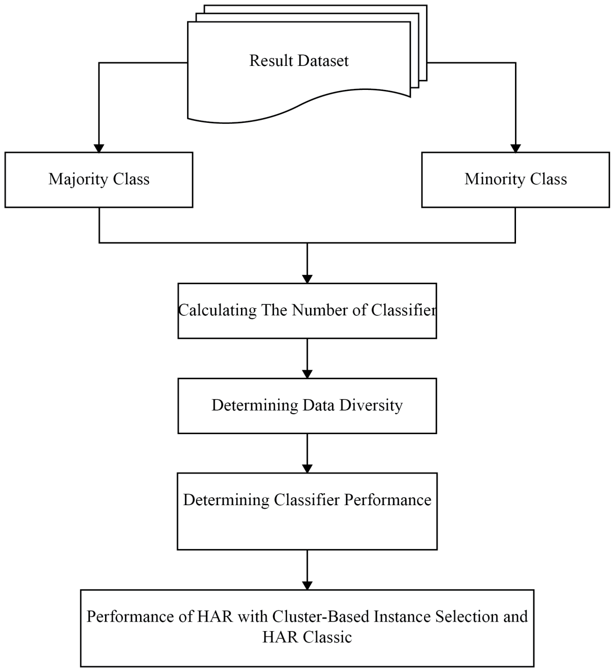

Fig. 3. Evaluation Stage

Based on Figure 3, it can be seen that the processing dataset will measure the number of classifiers, data diversity, and performance classifier. The results are expected to illustrate the performance of the results of handling the class imbalance between the results obtained by the Hybrid Approach Redefinition with Cluster-Based Instance Selection compare with classic Hybrid Approach Redefinition.

3.4. Under Sampling with Cluster-Based Instance Selection

The pseudocode of the Under Sampling process with Cluster-Based Instance Selection is as follows[13].

$Let\ S = Majority\ Class$

$Delete\ ncol(S)\ as\ the\ class\ label\ of\ S$

$Execute\ S\ to\ obtain\ the\ clusering\ list\ named\ AP$

$Allocate\ each\ record\ of\ S\ to\ Size\ (AP) Clusters$

$For\ (i = 1; i \leq size\ (AP); i++)$

{

  $For\ (j = 1; j \leq size\ (AP[i]); j++)$

  {

    $Value = AP[i][j]$

    $S[Value, ncol(S)] = i$

  }

}

$Performance\ Instance\ Selection\ Over\ S\ to\ Produce\ Suvsets\ S_{Noisy} and\ S_{NonNoisy}$

$Replace\ ncol(S)\ of\ all\ records\ in\ S_{nonnoisy} with\ the\ Majority\ Class$

$Let\ RR = nrow\ (S_{Noisy})/nrow(S)$

$Return\ S_{Noisy} and\ RR$

In the above pseudocode, it can be seen that if there are problems with imbalance class, based on the CBIS method, cluster analysis and division will be carried out into several clusters in the majority class. Where will be given a specific class label such as a class ID for each cluster that exists. Then the next step is to determine the instance of the majority class into the existing cluster. Then after that, an instance selection process will be carried out using the SMOTEBoost method. The process of this instance selection will produce an instance that has the best closeness with majority class. After going through the process, a reduced subset which will then undergo a noisy removal process produces S_Nonnoisy which is an instance selected as majority class.

3.5. Hybrid Approach Redefinition with Cluster-Based Instance Selection

The pseudocode of the Hybrid Approach Redefinition with Cluster-Based Instance Selection[20][13][21].

$Input\text{:}\ Total\ Size\ totalSize, Number\ of\ Majority\ S_N, Number\ of\ Minority\ S_P$

$totalSize \leftarrow |S|$

$S_N = \{(x_i, y_i) \in S | y_i = -1\}$

$S_P = \{(x_i, y_i) \in S | y_i = +1\}$

$majoritySize \leftarrow |S_N|$

$minoritySize \leftarrow |S_P|$

$\boldsymbol{Preprocessing\ Stage}$:

$Execute\ S_N\ to\ obtain\ the\ clusering\ list\ named\ AP$

$Allocate\ each\ record\ of\ S_N\ to\ Size\ (AP) Clusters$

$For\ (i = 1; i \leq size\ (AP); i++)$

{

```
  For (j = 1; j ≤ size (AP[i]); j++)
  {
    Value = AP[i][j]
    S[Value, ncol(S)] = i
  }
}
k = Number of Nearest Neighbors
numattrs = number of attributes
Sample[ ][ ]: Minority Class Sample
DMajority_Reduced = Array of Majority
DMinority = Array of Minority
For (i = 1; i ≤ majoritySize; i++)
{
  Compute k nearest neighbors
  Populate (N, i, nnarray)
}
While N ≠ 0 do
{
  for (i = 1; i ≤ numattrs; i++)
  {
    dif[i] = sample[nnarray[i][attr]-sample[i][attr]
  }
}
For (i = 1; i ≤ majoritySize; i++)
{
  Sort sample[i][attr]according to dif[i]
}
majoritySize_reduced = random number 1 to majoritySize
For (i = 1; i ≤ majoritySize ; i++)
{
  if i ≤ majoritySize_reduced
     DMajority_Reduced[i][attr] = sample[i][attr]
  else
     DMinority[i][attr] = sample[i][attr]
Combine DMajority_Reduced with D_Minority become D'
Processing Stage:
T = number of Iteration
for(i = 1; i ≤ T; i++)
{
  Classifying D'using B-SVM
```

```
Identifying Majority Class
Identifying Minority Class
While (!endOfMajorityClass) do
 {
   NewSVSets ← Deleting Noise of SVSets
   NewNSVSets ← Sampling NSVSets using CBIS
  }
While (!endOfMinorityClass) do
 {
   NewSVSets ← Deleting Noise of SVSets and Instance Selection using SMOTE
  }
}
Result Dataset
```

In the pseudocode, it can be seen that if the classification results indicate a class imbalance problem, the preprocessing stage begins with CBIS which performs cluster analysis where at this stage majority classes will be divided into a number of specific clusters and each existing instance is inserted into a particular cluster specific, so that each instance will be associated with a particular label class. The formation of the New Majority Class will be based on the existing class label information. The next step is to process an instance selection using SMOTE which will be done based on the generation of random numbers to move a number of instances to the majority class to minority class. This is based on the level of closeness of the existing instance to minority class. After this process is done, the results will be combined with the existing minority class to form a preprocessing dataset denoted by D '. The next process will go into the processing stage, which at this stage will involve the Biased Support Vector Machine method which will group both majority and minority classes into 2 (two) groups, namely: SV Sets and NSV Sets. First, noise cleaning will be done on the SV Sets Majority Class and NSV Minority Class Sets. The next step is NSV Sets for majority classes to undergo a CBIS sampling process, which is then followed by the process of instance selection on the NSV Minority Class by using SMOTEBoost. This process will produce a New Majority Class and New Minority Class, which will then be combined to form the Result dataset.

3.6. Data Diversity

Data diversity is intended to measure the performance of a classifier, especially in situations where misclassification occurs. Misclassification is an unavoidable thing in handling imbalance classes. However, on the other hand a small amount of misclassification is not necessarily good, because when faced with a situation where the number of instances in a minority class is very small, it is easy to group all instances into majority classes. Therefore, good diversity data shows that the misclassification does not only occur in one class and if there is misclassification it should be able to be covered by merging with another classifier[22][23].

Suppose that Z= $\{z_1, \ldots, z_n\}$ which is a dataset that is in the decision region $\Re^n$, so that $z_j \in \Re^n$ it is an instance involved in the classification problem. Then the output of the classifier D_i as a classifier paired comparison matrix (relationship pairwise classifier) can be seen in Table 1.

Table 1. Relationship Pairwise Classifier Matrix

	D_k Correct (1)	D_k Wrong (0)
D_i Correct (1)	N^{11}	N^{10}
D_i Wrong (0)	N^{01}	N^{10}

Diversity data can be calculated using Q-Statistics[24].

$$Q_{i,k} = \frac{N^{11}N^{00}-N^{01}N^{10}}{N^{11}N^{00}+N^{01}N^{10}} \tag{1}$$

3.7. Classifier Performance

On Binary Class issues, positive samples refer to minority class and negative samples refer to majority class. For the general classification results the classification results can be grouped into 4 (four) groups, namely: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) and can be presented in the Confusion Matrix as can be seen in Table 2 [25].

Table 2. Confusion Matrix for A Binary Class Problem

		Predicted (Classified) as	
		Positive	**Class Negative**
Actually (Really is)	**Positive Samples**	True Positive (TP)	False Negative (FN)
	Negative Samples	False Positive (FP)	True Negative (TN)

For the measurement of performance classifier based on the confusion matrix which can be seen in Table 2., the measurement is done based on several things as follows.

1. True Positive (TP) states the number of positive samples that are classified correctly as positive.
2. True Negative (TN) states the number of negative samples that are classified correctly as negative.
3. False Positive (FP) states the number of negative samples are classified incorrectly as positive.
4. False Negative (FN) states the amount of positive samples which are classified incorrectly as negative.

The classifier performance that can be measured based on the confusion matrix is as follows[25].

1. Sensitivity

Sensitivity states the ability of the classifier to correctly identify positive samples. The best value for sensitivity is 1 and the lowest value is 0. The sensitivity can be measured using Equation 2.

$$Sensitivity = \frac{TP}{TP+FN} \tag{2}$$

2. Specificity

Specificity states the ability of the classifier to correctly identify the negative sample. The best value for specificity is 1 and the lowest value is 0. The specificity can be measured using Equation 3.

$$Specificity = \frac{TN}{TN+FP} \tag{3}$$

3. F-Measure

The term F-Measure usually refers to the harmonious average value between Precision and Recall. Precision states how well the classifier avoids the misclassification of the negative class as a positive class and recall states how well the classifier classifies the positive class.

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F\text{-}Measure = \frac{2RP}{R+P} \tag{6}$$

4. G-Means

G-Mean states on the ability of the classifier to achieve the balance of the accuracy of the classification on positive samples and negative samples.

$$G\text{-}Mean = \sqrt{Sensitivity\ .\ Specificity} \quad (7)$$

4. Result and Discussion

4.1. Dataset Description

The dataset used in this research are Page-Blocks, Vowel, and Vehicle1. The description about dataset can be seen in Table 3.

Table 3. Dataset Description

Dataset	#Ex	#Atts	(%Min;%Max)	IR
Page-Blocks	5472	10	(10.23,89.77)	8.77
Vowel	988	13	(9.01,90.99)	10.10
Vehicle1	846	18	(28.37,71.63)	2.52

In Table 3 it can be seen that the dataset selected is a dataset with a large number of attributes. In general, the datasets represent the number of instances that vary: small, medium and large and also with small, medium and large imbalance ratios.

4.2. Testing

Tests will be conducted to obtain a performance picture of the Hybrid Approach Redefinition method with Cluster-Based Instance Selection and Classic Hybrid Approach Redefinition, especially in terms of the number of classifiers, diversity data, and also the performance classifier. The test will be carried out 10 times for each method. The test results for the number of classifier and diversity data can be seen in Table 4.

Table 4. Testing Result for Number of Classifier and Data Diversity of Each Method

Dataset	Hybrid Approach Redefinition		Hybrid Approach Redefinition with Cluster-Based Instance Selection	
	Number of Classifier	Data Diversity	Number of Classifier	Data Diversity
Page-Blocks	517.1	0.916	512	0.878
Vowel	214.1	0.264	207.2	0.259
Vehicle1	207.1	0.615	206	0.595

Based on Table 4, it can be seen that for the number of classifier and data diversity, Hybrid Approach Redefinition with Cluster-Based Instance Selection can provide better results compared to the classic Hybrid Approach Redefinition. However, if it is seen that diversity data for Page-Blocks in both methods is still not good. Classic Hybrid Approach Redefinition gives results of 0.916 and Hybrid Approach Redefinition with Cluster-Based Instance Selection giving a result of 0.878. Even though Hybrid Approach Redefinition with Cluster-Based Instance Selection still provides better results compared to Hybrid Approach Redefinition, the results obtained should be better. However, for a small number of instances the data diversity results are quite good.

Comment [Office1]: No empirical relations used to validate techniques mention.

The measurement results for the performance classifier measured based on sensitivity, specificity, F-Measure, and G-Mean can be seen in Table 5.

Table 5. Testing Result for Sensitivity, Specificity, F-Measure, and G-Mean of Each Method

Dataset	Hybrid Approach Redefinition				Hybrid Approach Redefinition with Cluster-Based Instance Selection			
	Sensitivity	Specificity	F-Measure	G-Mean	Sensitivity	Specificity	F-Measure	G-Mean
Page-Blocks	0.521	0.918	0.542	0.691	0.542	0.921	0.567	0.71
Vowel	0.537	0.763	0.632	0.64	0.529	0.771	0.651	0.64
Vehicle1	0.467	0.496	0.64	0.481	0.471	0.492	0.701	0.481

In Table 5 it can be seen that in general the results of performance classifier measurements show that Hybrid Approach Redefinition with Cluster-Based Instance Selection gives better results compared to classic Hybrid Approach Redefinition. Measurements for sensitivity, specificity, and F-Measure tend to show Hybrid Approach Redefinition with Cluster-Based Instance Selection to give better results. Whereas for the G-Mean the results obtained are not much different and for the Dataset and Vehicle1 the results given by the two methods are the same. The measurement for G-Mean given by both methods is good and this means that the balance of predictive accuracy for majority classes and minority classes is quite good.

4.3. Discussion

Based on a series of tests that in terms of the number of classifiers that Hybrid Approach Redefinition with Cluster-Based Instance Selection can reduce the number of classifiers, but not too much different. This is because classic Hybrid Approach Redefinition in general has been able to overcome the problem of the imbalance class with a very good number of classifiers. The sampling process in classic Hybrid Approach Redefinition that uses Random Under Sampling gives only slightly worse results than under sampling using Cluster-Based Instance Selection.

However, what needs to be paid attention to is the data diversity, there is a tendency that the two methods have not been very effective in the dataset with a large number of instances. This means that there is a tendency for misclassification to occur in only one classifier group. This might be overcome by selecting a more appropriate instance selection method. Where based on research from Soleymani et al[26]. that SMOTEBoost does have a tendency to have weaknesses in providing good data diversity.

The measurement results for sensitivity, specificity, F-Measure, and G-Mean given are very good. So there is no concern that there is a high number of misclassification in the minority class and majority class. A good accuracy balance is shown in the G-Mean value which tends to be the same in the Hybrid Approach Redefinition with Cluster-Based Instance Selection and Classic Hybrid Approach Redefinition.

Through the results of this study it can be obtained that the Hybrid Approach Redefinition with Cluster-Based Instance Selection provides better results compared to the classic Hybrid Approach Redefinition, both for the number of classifiers, diversity data, and performance classifiers. Future research is expected to focus on the problem of data diversity, especially for datasets with large number of instances.

Comment [Office2]: It would be better if some more comprehensive evaluations are included.

5. Conclusion

The study implemented Hybrid Approach Redefinition with Cluster-Based Instance Selection in handling class imbalance problem. The results showed that Hybrid Approach Redefinition with cluster-based instance selection gave better results on the number of classifiers, data diversity, and performance classifiers compared to classic Hybrid Approach Redefinition. This research discusses handling class imbalance for two-class imbalance problems and future research can develop this method so that it can handle Multi-class imbalance problems.

Comment [Office3]: Where are the conclusions and future work? Please include. ... [1]

References

[1] C. K. Maurya, D. Toshniwal, and G. V. Venkoparao, "Online anomaly detection via class-imbalance learning," presented at the Eighth International Conference on Contemporary Computing (IC3), 2015, pp. 30–35.

[2] N. V. Chawla, "Data Mining for Imbalanced Datasets: An Overview | SpringerLink," in *Data Mining and Knowledge Discovery Handbook*, Springer, 2005, pp. 853–867.

[3] Z. P. Agusta and A. Adiwijaya, "Modified balanced random forest for improving imbalanced data prediction," *International Journal of Advances in Intelligent Informatics*, vol. 5, no. 1, pp. 58–65, 2019.

[4] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, Nov. 2016.

[5] C. K. Maurya and D. Toshniwal, "Large-Scale Distributed Sparse Class-Imbalance Learning," *Information Sciences*, vol. 456, pp. 1–11, 2018.

[6] 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.

[7] C. Jian, J. Gao, and Y. Ao, "A new sampling method for classifying imbalanced data based on support vector machine ensemble," *Neurocomputing*, vol. 193, pp. 115–122, Jun. 2016.

[8] M.-J. Kim, D.-K. Kang, and H. B. Kim, "Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1074–1082, Feb. 2015.

[9] P. Yang, P. D. Yoo, J. Fernando, B. B. Zhou, Z. Zhang, and A. Y. Zomaya, "Sample Subset Optimization Techniques for Imbalanced and Ensemble Learning Problems in Bioinformatics Applications," *IEEE Transaction on Cybernetics*, vol. 44, no. 3, 2014.

[10] Hartono, E. Ongko, O. S. Sitompul, Tulus, E. B. Nababan, and D. Abdullah, "Hybrid Approach Redefinition (HAR) Method with Loss Factors in Handling Class Imbalance Problem," *2018 International Symposium on Advanced Intelligent Informatics (SAIN)*, pp. 56–61, 2018.

[11] H. Hartono, O. S. Sitompul, T. Tulus, and E. B. Nababan, "Biased support vector machine and weighted-smote in handling class imbalance problem," *International Journal of Advances in Intelligent Informatics*, vol. 4, no. 1, pp. 21–27, Apr. 2018.

[12] I. C. Irsan and M. L. Khodra, "Hierarchical multi-label news article classification with distributed semantic model based features," *International Journal of Advances in Intelligent Informatics*, vol. 5, no. 1, pp. 40–47, 2019.

[13] C.-F. Tsai, W.-C. Lin, Y.-H. Hu, and G.-T. Yao, "Under-sampling class imbalanced datasets by combining clustering analysis and instance selection," *Information Sciences*, vol. 477, pp. 47–54, 2018.

[14] 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.

[15] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[16] D. R. Wilson and T. R. Martinez, "Reduction Techniques for Instance-Based Learning Algorithms," *Machine Learning*, vol. 38, no. 3, pp. 257–286, Mar. 2000.

[17] V. Vigneron and H. Chen, "A multi-scale seriation algorithm for clustering sparse imbalanced data: application to spike sorting," *Pattern Anal Applic*, vol. 19, no. 4, pp. 885–903, Nov. 2016.

[18] S. Garcia, J. Derrac, J. Cano, and F. Herrera, "Prototype Selection for Nearest Neighbor Classification: Taxonomy and Empirical Study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 417–435, Mar. 2012.

[19] 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.

[20] Hartono, O. S. Sitompul, E. B. Nababan, Tulus, D. Abdullah, and A. S. Ahmar, "A New Diversity Technique for Imbalance Learning Ensembles," *International Journal of Engineering & Technology*, vol. 7, no. 2, pp. 478–483, Apr. 2018.

[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[22] J. F. Díez-Pastor, J. J. Rodríguez, C. I. García-Osorio, and L. I. Kuncheva, "Diversity techniques improve the performance of the best imbalance learning ensembles," *Information Sciences*, vol. 325, pp. 98–117, Dec. 2015.

[23] L. I. Kuncheva, *Combining Pattern Classifiers*. John Wiley & Sons, 2004.

[24] G. U. Yule, "On the Association of Attributes in Statistics: With Illustrations from the Material of the Childhood Society, &c," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 194, pp. 257–319, 1900.

[25] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A Review," *International Journal of Advances in Soft Computing and Its Application*, vol. 7, no. 3, pp. 176–204, 2015.

[26] R. Soleymani, E. Granger, and G. Fumera, "Progressive boosting for class imbalance and its application to face re-identification," *Expert Systems with Applications*, vol. 101, pp. 271–291, Jul. 2018.

Page 11: [1] Commented **Microsoft Office User** **2/24/24 9:56 PM**

Where are the conclusions and future work? Please include.

Gmail

Hartono Ibbi <hartonoibbi@gmail.com>

[SAIN'19] Paper 1570568696 has been uploaded

1 pesan

sain@ijain.org <sain@ijain.org> 11 Agustus 2019 pukul 00.00
Balas Ke: sain@ijain.org
Cc: Andri Pranolo <andri.pranolo@tif.uad.ac.id>, Adhi Prahara <adhi.prahara@tif.uad.ac.id>, Hartono Hartono <hartonoibbi@gmail.com>, Erianto Ongko <eriantoongko@gmail.com>, Dahlan Adbullah <dahlan.unimal@gmail.com>

Dear Mr. Hartono Hartono:

Thank you for uploading your paper 1570568696 ('Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem') to 2019 2nd International Symposium on Advanced Intelligent Informatics (SAIN). The paper is of type application/msword and has a length of 5429760 bytes.

You can modify your paper at https://edas.info/showPaper.php?m=1570568696 and see all your submissions at https://edas.info/r26041 using the EDAS identifier hartonoibbi@gmail.com

Regards,
The Technical program chairs

Hasil Review 9 November 2021

Editor
2021-11-09 12:29 PM

Subject: [IJAIN] Editor Decision for Paper entitled (Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem)

DELETE

Dear Mr. Hartono

We have reached a decision regarding your submission to the International Journal of Advances in Intelligent Informatics, "Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem".

Our decision is: Accept with Minor Revisions

Please kindly submit the revision TWO WEEKS after receiving this notification, and follow the instructions carefully,

1. Do the corrections with track changes.

2. We required 3 files as feedback, a) File with track changes corrections;
b) A file without track changes (Final copy/clean copy); c) Table of
correction as a response to editors/ Reviewers comments. Upload all files in
*.ZIP extension file.

3. Follow IJAIN Author guidelines at http://ijain.org/index.php/IJAIN/about/submissions#authorGuidelines

Please NOTED that if the author(s) not follow the feedback instruction and submit the revisions at the time, it would be editor(s) reasons to DECLINE your submission.

Should you have any queries please do not hesitate to contact us by email. We look forward to hearing from you.

Regards,

Andri Pranolo
(Editor-in-Chief)
info@ijain.org

Reviewer M:

Significance:
- How important is the work reported? Does it attack an important/difficult problem (as opposed to a peripheral/simple one)?
- Does the approach offered advance the state of the art?
- Does it involve or synthesize ideas, methods, approaches from multiple disciplines?
- Does it have interesting implications for multiple disciplines?:
Good

Originality: - Is this a new issue? Is this a novel approach to an issue? - Is this a novel combination of familiar ideas/techniques/methods/approaches? - Does the paper point out differences from related research? - Does the paper properly situate itself with respect to previous work?:
Good

Quality: - Is the paper technically sound? How are its claims backed up? - Does it carefully evaluate the strengths and limitations of its contribution?:
Good

Clarity: - Is the paper clearly written? Does it motivate the research? Does it describe clearly the methods employed (e.g., experimental procedures, algorithms, analytical tools), if any? - Are the results, if any, described and evaluated thoroughly? - Is the paper organized in a sensible and logical fashion?:
Good

Relevance:
- Is the paper closely related to the theme of the journal (broadly conceived)?
- Is the content interesting enough to a broad audience?
- Is the paper readable in a multi-disciplinary context?:
Good

Technical (1): Structure of the paper:
Good

Technical (2): Standard of English:
Good

Technical (3): Appropriateness of abstract as a description of the paper:
Good

Technical (4): Use and number of keywords/key phrases:
Good

Technical (5): Relevance and clarity of drawings, graphs and tables:
Fair

Technical (6): Discussion and conclusions:
Good

Technical (7): Reference list, adequate and correctly cited:
Good

Explanations for the above ratings and other general comments on major issues:

Comments on the minor details of the article:
1. Abstract must have 2-3 lines about implications of the study.
2. Author must cite more papers in related work.
3. Algorithm is written well but author must use legitimate style of writing algorithm. please see the springer and IEEE journal paper for writing algorithm.
4. If possible author must make use of graphs to understand the paper better to the readers.

Reviewer N:

Significance:
- How important is the work reported? Does it attack an important/difficult problem (as opposed to a peripheral/simple one)?
- Does the approach offered advance the state of the art?
- Does it involve or synthesize ideas, methods, approaches from multiple disciplines?
- Does it have interesting implications for multiple disciplines?:
Good

Originality: - Is this a new issue? Is this a novel approach to an issue? - Is this a novel combination of familiar ideas/techniques/methods/approaches? - Does the paper point out differences from related research? - Does the paper properly situate itself with respect to previous work?:
Fair

Quality: - Is the paper technically sound? How are its claims backed up? - Does it carefully evaluate the strengths and limitations of its contribution?:
Good

Clarity: - Is the paper clearly written? Does it motivate the research? Does it describe clearly the methods employed (e.g., experimental procedures, algorithms, analytical tools), if any? - Are the results, if any, described and evaluated thoroughly? - Is the paper organized in a sensible and logical fashion?:
Fair

Relevance:
- Is the paper closely related to the theme of the journal (broadly conceived)?
- Is the content interesting enough to a broad audience?
- Is the paper readable in a multi-disciplinary context?:
Good

Technical (1): Structure of the paper:
Good

Technical (2): Standard of English:
Fair

Technical (3): Appropriateness of abstract as a description of the paper:
Good

Technical (4): Use and number of keywords/key phrases:
Good

Technical (5): Relevance and clarity of drawings, graphs and tables:
Good

Technical (6): Discussion and conclusions:
Fair

Technical (7): Reference list, adequate and correctly cited:
Good

Explanations for the above ratings and other general comments on major issues:
The paper addresses an important topic which is handling class imbalance problem. The paper is well structured, written and compared.
Check the table on page 8.
I suggest using figure or listing for pseudocode. Also, rephrase the discussion.

Comments on the minor details of the article:
Explain equation 1 and 6.

Reviewer T:

Significance:
- How important is the work reported? Does it attack an important/difficult problem (as opposed to a peripheral/simple one)?
- Does the approach offered advance the state of the art?
- Does it involve or synthesize ideas, methods, approaches from multiple disciplines?
- Does it have interesting implications for multiple disciplines?:
Fair

Originality: - Is this a new issue? Is this a novel approach to an issue? - Is this a novel combination of familiar ideas/techniques/methods/approaches? - Does the paper point out differences from related research? - Does the paper properly situate itself with respect to previous work?:
Fair

Quality: - Is the paper technically sound? How are its claims backed up? - Does it carefully evaluate the strengths and limitations of its contribution?:
Fair

Clarity: - Is the paper clearly written? Does it motivate the research? Does it describe clearly the methods employed (e.g., experimental procedures, algorithms, analytical tools), if any? - Are the results, if any, described and evaluated thoroughly? - Is the paper organized in a sensible and logical fashion?:
Fair

Relevance:
- Is the paper closely related to the theme of the journal (broadly conceived)?
- Is the content interesting enough to a broad audience?
- Is the paper readable in a multi-disciplinary context?:
Fair

Technical (1): Structure of the paper:
Good

Technical (2): Standard of English:
Poor

Technical (3): Appropriateness of abstract as a description of the paper:
Fair

Technical (4): Use and number of keywords/key phrases:
Good

Technical (5): Relevance and clarity of drawings, graphs and tables:
Fair

Technical (6): Discussion and conclusions:
Fair

Technical (7): Reference list, adequate and correctly cited:
Fair

Explanations for the above ratings and other general comments on major issues:
1. Specify what authors mean by "number of classifiers." A reference [13] is given for the same; However, it does not specify anything about the " number of classifiers."
2. Write expanded form of IR.
3. Specify the train-test partitioning, i.e., whether a k-fold or a random train test partitioning is used.
a. If k-fold partitioning is used, specify the value of "k."
b. If random partitioning is used, specify the train-test size.
4. Cite some recent articles, i.e., of the year 2021, for example
N. K. Mishra and P. K. Singh, "Feature construction and smote-based imbalance handling for multi-label learning," Information Sciences, 563, Pages 342-357, 2021.
5. Explain the similarity and differences of the proposed method with the reference [17].
6. If possible, compare the results with [17].
There are various linguistic errors. Carefully proofread the whole paper for better readability and grammatical correctness.

Comments on the minor details of the article:
1. Specify what authors mean by "number of classifiers." A reference [13] is given for the same; However, it does not specify anything about the " number of classifiers."
2. Write expanded form of IR.
3. Specify the train-test partitioning, i.e., whether a k-fold or a random train test partitioning is used.
a. If k-fold partitioning is used, specify the value of "k."
b. If random partitioning is used, specify the train-test size.
4. Cite some recent articles, i.e., of the year 2021, for example
N. K. Mishra and P. K. Singh, "Feature construction and smote-based imbalance handling for multi-label learning," Information Sciences, 563, Pages 342-357, 2021.
5. Explain the similarity and differences of the proposed method with the reference [17].
6. If possible, compare the results with [17].
There are various linguistic errors. Carefully proofread the whole paper for better readability and grammatical correctness.

International Journal of Advances in Intelligent Informatics (IJAIN)
ISSN 2442-6571 (print) | 2548-3161 (online)

SCOPUS Indexed Journal | CiteScore 2019 = 1.4 |
CiteScoreTracker 2020 = 1.9 (Last updated on 2 March, 2021)
SCOPUS: https://www.scopus.com/sourceid/21100890645

SJR 2019 = 0.17
SJR: https://www.scimagojr.com/journalsearch.php?q=21100890645&tip=sid&exact=no
----
Website: http://ijain.org/index.php/IJAIN
Contact: ijain@uad.ac.id, info@ijain.org

Editor
2021-11-22 10:08 AM

Subject: [IJAIN] Editor Final Decision DELETE

Mr. Hartono -:

We have reached a decision regarding your submission to International Journal of Advances in Intelligent Informatics, "Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem".

Our decision is to: Accept Submission

Please keep attention for the copy editing and proofreading process which are final publicity process on IJAIN Journal. Your paper is scheduled to be published in the upcoming issue after we finished those process.

Regards,

Andri Pranolo
(Editor-in-Chief

International Journal of Advances in Intelligent Informatics (IJAIN)
ISSN 2442-6571 (print) | 2548-3161 (online)

SCOPUS Indexed Journal | CiteScore 2019 = 1.4 |
CiteScoreTracker 2020 = 1.9 (Last updated on 2 March, 2021)
SCOPUS: https://www.scopus.com/sourceid/21100890645

SJR 2019 = 0.17
SJR: https://www.scimagojr.com/journalsearch.php?q=21100890645&tip=sid&exact=no
----
Website: http://ijain.org/index.php/IJAIN
Contact: ijain@uad.ac.id, info@ijain.org

TABLE OF CORRECTIONS

1. Correction according to comments from Reviewer M

No	Comments	Actions
1	Abstract must have 2-3 lines about implications of the study.	The implications of the study have been clarified by adding an explanation in the 8th sentence of the abstract
2	Author must cite more papers in related work	Citations of several papers have been added in the related works section. This can be seen in reference [25-28]
3	Algorithm is written well but author must use legitimate style of writing algorithm. please see the springer and IEEE journal paper for writing algorithm.	Algorithm writing is in accordance with IEEE style
4	If possible author must make use of graphs to understand the paper better to the readers.	Figure 1 has been added which explains the stages of the research

2. Correction according to comments from Reviewer N

No	Comments	Actions
1	Explain equation 1 and 6.	All equations in this study have been added with explanations.

3. Correction according to comments from Reviewer T

No	Comments	Actions
1	Specify what authors mean by "number of classifiers." A reference [13] is given for the same; However, it does not specify anything about the " number of classifiers."	An explanation of the number of classifiers has been added in the 2nd Paragraph in related works
2	Write expanded form of IR	The Imbalance Ratio (IR) has been explained in Section 3.6
3	Specify the train-test partitioning, i.e., whether a k-fold or a random train test partitioning is used.	K Fold-Cross Validation has been explained in section 4.2 testing
4	Cite some recent articles, i.e., of the year 2021, for example N. K. Mishra and P. K. Singh, "Feature construction and smote-based imbalance handling for multi-label learning," Information Sciences, 563, Pages 342-357, 2021.	References in the form of articles for 2021 have been added and can be seen in reference numbers [20], [21], [25], and [26].
5	Explain the similarity and differences of the proposed method with the reference [17]	Differences in research are not written in the article section because research [17] focuses on accuracy while this study focuses on number of classifiers, data diversity, and performance classifiers.

REVISI SESUAI MASUKAN DARI REVIEWER IJAIN

Gmail

Hartono Ibbi <hartonoibbi@gmail.com>

Revision for Paper entitled (Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem)

1 pesan

Hartono Ibbi <hartonoibbi@gmail.com> 22 November 2021 pukul 09.04

Kepada: Andri Pranolo <info@ijain.org>

Dear Mr. Andri Pranolo,
Good morning. Attached is the file for the revision of the IJAIN article with the title "Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem" which contains the following files.
a) File with track changes corrections;
b) A file without track changes (Final copy/clean copy);
c) Table of correction as a response to editors/ Reviewers comments

We understand that there may still be deficiencies in our revision. Please do not hesitate to let us know if there is anything further we need to improve.

Thank you for the help and comments given to us.

Best Regards,

Hartono

IJAIN Revision Files.zip
748K

International Journal of Advances in Intelligent Informatics
ISSN 2442-6571 (print) 2548-3161 (online) | available at http://ijain.org/index.php/IJAIN
Member of Crossref
Accreditated "1st" grade (SINTA 1) by Ministry of Research, Technology and Higher Education (RistekDikti) of the Republic of Indonesia, Decree No. 30/E/KPT/2018, October 24, 2018
Indexed in Scopus, sinta, ASEAN Citation Index, EBSCOhost, Academic Search, ProQuest
Published by UAD 1960

#515 Review

SUMMARY **REVIEW** EDITING

Submission

Authors	Hartono Hartono, Erianto Ongko, Dahlan Abdullah
Title	Hybrid approach redefinition with cluster-based instance selection in handling class imbalance problem
Section	Articles
Editor	Rafał Dreżewski (Review) Mohammad Syafrullah (Review)

Peer Review

Round 1

Review Version	515-1556-1-RV.DOC 2020-06-24
Initiated	2020-06-24
Last modified	2021-09-27
Uploaded file	Reviewer N 515-2115-1-RV.PDF 2021-09-14

Editor Decision

Decision	Accept Submission 2021-11-22
Notify Editor	Editor/Author Email Record 2021-11-22
Editor Version	None
Author Version	515-1907-2-ED.DOCX 2021-10-31 DELETE 515-1907-3-ED.ZIP 2021-11-22 DELETE
Upload Author Version	Choose File No file chosen Upload

QUICK MENU
Editorial Team
Reviewer
Focus and Scope
Author Guidelines
Publication Ethics
Open Access Policy
Peer Review Process
Online Submission
Author(s) Fee
Contact

USER
You are logged in as...
hartono
- My Profile
- Log Out

TEMPLATE

JOURNAL CONTENT
Search
Search Scope
All
Search
Browse
- By Issue
- By Author
- By Title

TOOLS
MENDELEY
iThenticate
grammarly

CURRENT INDEXING

CURRENT INDEXING

More Abstracting & Indexing...

REQUEST INDEXING

- **SCOPUS (ACCEPTED)**
  - Submission Received: **July 26, 2017**
  - Submission Accepted: **June 5, 2018**
  - **SCOPUS CiteScore Tracker 2020**
- **Web of Science**
  - Latest submission: June 11, 2018
  - **Web of Science** Citation Analysis
- **IET INSPEC**
  - Added to review: **June 5, 2018**
- **COMPENDEX**
  - Submission: **October 15, 2017**
  - Resubmission: **September 27, 2020**

CROSSREF

INFORMATION

- For Readers
- For Authors

ISSN BARCODE

9 772548 316004

ISSN Online

9 772442 657005

ISSN Print

KEYWORDS

CNN COVID-19 Class Imbalance Classification Clustering Computer vision Deep Learning Deep learning Effort estimation Feature selection GPU LSTM Machine Learning Machine learning Neural network Ontology Optimization SVM Scheduling Support Vector Machines (SVM) Transfer Learning