

PAPER NAME

3 Jurnal Nasional Hartono.pdf

AUTHOR

Hartono

WORD COUNT

1240 Words

CHARACTER COUNT

7736 Characters

PAGE COUNT

4 Pages

FILE SIZE

622.3KB

SUBMISSION DATE

Feb 24, 2024 1:11 PM GMT+7

REPORT DATE

Feb 24, 2024 1:11 PM GMT+7

● 11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 10% Publications database
- Crossref Posted Content database
- Crossref database
- 8% Submitted Works database

● Excluded from Similarity Report

- Internet database
- Bibliographic material

Metode Hybrid Approach-RCSMOTE di dalam Penanganan Class Imbalance

Hartono

1 Program Studi Magister Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama

e-Mail: hartonoibbi@gmail.com

Abstrak

Class Imbalance merupakan salah satu permasalahan di dalam klasifikasi. *Class imbalance* adalah permasalahan dimana jumlah *instance* salah satu *class* jauh lebih besar dibandingkan dengan *class* lainnya. *Class* dengan jumlah *instance* yang besar dinamakan *majority class* dan *class* dengan jumlah *instance* yang lebih kecil dinamakan sebagai *minority class*. Adanya *class imbalance* ini dapat menyebabkan akurasi yang rendah dan gagal diperolehnya pola yang menarik pada *minority class*. Salah satu metode yang baik di dalam penanganan *class imbalance* adalah *Hybrid Approach*. Metode *Hybrid Approach* merupakan penggabungan dari *data level* dan *algorithm level*. *Data level* menggunakan proses *oversampling* dan *undersampling*, tetapi yang umum digunakan adalah *oversampling* dengan menggunakan SMOTE. Permasalahan *oversampling* sering kali menimbulkan *overfitting*. Oleh karena itu, metode *Range-Controlled Synthetic Minority Over-Sampling Technique* (RCSMOTE) yang bertujuan untuk menyempurnakan proses *oversampling* pada SMOTE. Penelitian ini akan menggunakan metode RCSMOTE pada *Hybrid Approach*. Adanya metode *Hybrid Approach-RCSMOTE* diharapkan dapat meningkatkan kinerja pada penanganan *class imbalance*. Hasil penelitian menunjukkan bahwa *Hybrid Approach-RCSMOTE* dapat memberikan hasil penanganan *class imbalance* dalam bentuk jumlah misklasifikasi dan *performance classifier* yang lebih baik.

Kata Kunci: *Class Imbalance, Hybrid Approach, SMOTE, RCSMOTE*

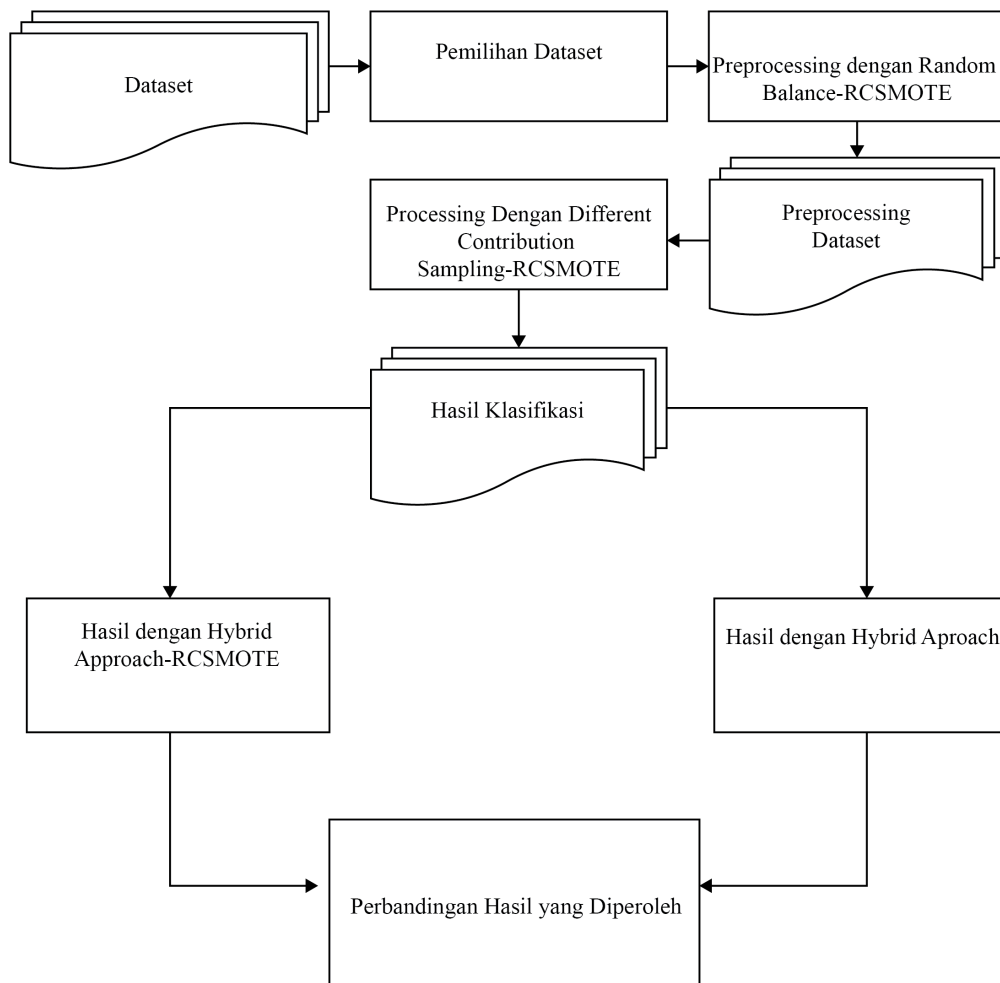
1. PENDAHULUAN

Permasalahan *class imbalance* sering dijumpai dalam kehidupan sehari-hari (Yuan et al., 2018). Permasalahan *class imbalance* ditunjukkan dengan adanya *class* dengan jumlah *instance* yang jauh lebih besar dibandingkan dengan *class* lain sebagai contoh data medis cenderung mengalami *class imbalance* dimana data dengan jumlah kasus positif jauh lebih kecil dibandingkan dengan data jumlah kasus negatif (Wang & Cheng, 2021). Permasalahan ini dapat menimbulkan masalah karena akurasi prediksi akan lebih baik pada *class* dengan jumlah *instance* yang lebih besar dibandingkan dengan akurasi pada *class* dengan jumlah *instance* yang lebih kecil (Mortaz, 2020). Hal lain yang menjadi masalah adalah bahwa jika terjadi *class imbalance* maka pola menarik pada *minority class* tidak akan dapat diperoleh (Li et al., 2021).

Penanganan *class imbalance* secara umum dapat dibagi menjadi 3 (tiga) pendekatan, yakni: *Data Level*, *Algorithm Level*, dan *Hybrid Approach* (Lin et al., 2017). *Data level* menggunakan proses *oversampling* dan *undersampling*, tetapi yang umum digunakan adalah *oversampling* dengan menggunakan SMOTE. Permasalahan *oversampling* sering kali menimbulkan *overfitting*. Oleh karena itu, metode *Range-Controlled Synthetic Minority Over-Sampling Technique* (RCSMOTE) yang bertujuan untuk menyempurnakan proses *oversampling* pada SMOTE (Soltanzadeh & Hashemzadeh, 2021). Adapun *Hybrid Approach* menggabungkan pendekatan *data level* dan *algorithm level*. Penelitian ini akan menggunakan metode RCSMOTE pada *Hybrid Approach*. Adanya metode *Hybrid Approach-RCSMOTE* diharapkan dapat meningkatkan kinerja pada penanganan *class imbalance*.

2. METODE

7 Adapun tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pada Gambar 1 dapat dilihat bahwa terdapat perbedaan antara *Hybrid Approach-RCSMOTE* dengan *Hybrid Approach*. Pada *Hybrid Approach*, tahapan *preprocessing* dilakukan dengan menggunakan *Random Balance Ensemble Method* yang menggunakan *SMOTE* sedangkan pada metode yang diusulkan menggunakan *Range-Controlled Synthetic Minority Over-Sampling Technique (RCSMOTE)*. Demikian juga halnya dengan tahapan *processing*, pada *Hybrid Approach* menggunakan *Different Contribution Sampling-SMOTE* dan pada *Hybrid Approach-RCSMOTE* menggunakan *Different Contribution Sampling-RCSMOTE*.

3. HASIL DAN PEMBAHASAN

3.1. Hybrid Approach

Adapun *pseudocode* dari *Hybrid Approach* adalah sebagai berikut (Galar et al., 2012).

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

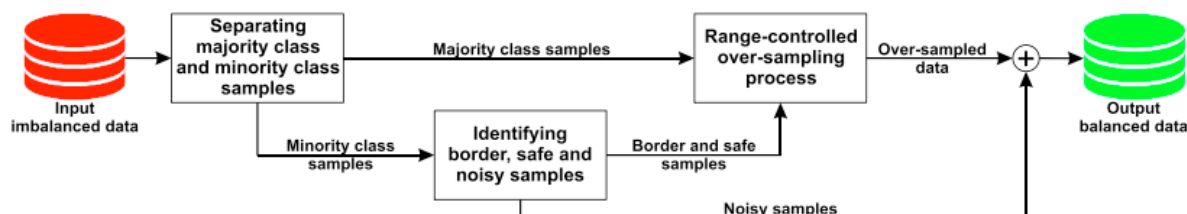
Step 3 For $i = 1$ to n

Apply processing using bagging, boosting or sampling

End For

3.2. RCSMOTE

Adapun tahapan dari RCSMOTE dapat dilihat pada Gambar 2 (Soltanzadeh & Hashemzadeh, 2021).



Gambar 2. Tahapan RCSMOTE

Berdasarkan pada Gambar 2 dapat dilihat bahwa perbedaan utama dengan SMOTE adalah terdapat proses *Range-Controlled* yang dilakukan berdasarkan nilai maksimum P_Max dan nilai minimum N_Min dari tiap atribut. Hal ini dapat dilakukan dengan menggunakan Persamaan 1.

$$Range = \frac{(P_{Max} + N_{Min})}{2} \tag{1}$$

3.3. Dataset

Adapun dataset yang digunakan di dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Dataset yang Digunakan

No	Dataset	Jumlah Instance	Jumlah Atribut	(%Min;%Max)	Imbalanced Ratio (IR)
1	Iris	150	4	(33.33,66.67)	2
2	Glass	214	9	(32.71,67.29)	2.06
3	Yeast1	1484	8	(28.91,71.09)	2.46
4	Vehicle1	846	18	(28.37,71.63)	2.52

3.4. Classifier Performance

Adapun pengukuran *classifier performance* dilakukan dengan sejumlah parameter yang diperoleh berdasarkan *confusion matrix* yang dapat dilihat pada Tabel 2.

Tabel 2. Confusion Matrix

	Diklasifikasikan sebagai Positive	Diklasifikasikan sebagai Negative
Positive Samples	True Positive (TP)	False Negative (FN)
Negative Samples	False Positive (FP)	True Negative (TN)

$$True\ Negative\ Rate\ (T\ Nrate) = \frac{TN}{TN + FP} \tag{2}$$

$$False\ Negative\ Rate\ (F\ Nrate) = \frac{FN}{TP + FN} \tag{3}$$

$$Positive\ Predictive\ Value\ (PPValue) = \frac{TP}{TP + FP} \tag{4}$$

$$Negative\ Predictive\ Value\ (NPValue) = \frac{TN}{TN + FN} \tag{5}$$

$$\underline{Recall} = TPrate = \frac{TP}{TP + FN} \quad (6)$$

$$\underline{Precision} = PPValue = \frac{TP}{TP + FP} \quad (7)$$

3.5. Hasil

Adapun hasil pengukuran akurasi dan *classifier performance* dengan 20 kali pengujian dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengukuran Akurasi dan *Classifier Performance*

Dataset	Hybrid Approach-RCSMOTE			Hybrid Approach		
	Misklasifikasi	Precision	Recall	Misklasifikasi	Precision	Recall
Iris	10.9	0.91	0.92	12.5	0.87	0.89
Glass	17.9	0.90	0.87	18.3	0.85	0.83
Yeast1	18.8	0.87	0.83	23.5	0.82	0.81
Vehicle1	20.6	0.88	0.86	26.7	0.76	0.79

Berdasarkan pada Tabel 3 dapat dilihat bahwa hasil yang diperoleh dengan *Hybrid Approach-RCSMOTE* lebih baik jika dibandingkan dengan *Hybrid Approach* yang menggunakan SMOTE.

4. KESIMPULAN

4.1. Kesimpulan

Adapun kesimpulan dari penelitian ini adalah bahwa hasil yang diperoleh oleh metode *Hybrid Approach-RCSMOTE* lebih baik dibandingkan dengan *Hybrid Approach*. Hasil yang lebih baik ini diperoleh melalui proses perbaikan pada *oversampling* dimana metode SMOTE diganti dengan RCSMOTE yang menunjukkan kinerja yang lebih baik di dalam proses *oversampling*.

4.2. Saran

Adapun saran yang dapat diberikan adalah bahwa hendaknya proses pengujian dapat menambah *dataset* dan juga parameter pengukuran yang digunakan dapat ditambah.

DAFTAR PUSTAKA

- [1] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2012). 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)*, 42(4), 463–484. <https://doi.org/10.1109/TSMCC.2011.2161285>
- [2] Li, J., Zhu, Q., Wu, Q., & Fan, Z. (2021). A novel oversampling technique for class-imbalanced learning based on SMOTE and natural neighbors. *Information Sciences*, 565, 438–455. <https://doi.org/10.1016/j.ins.2021.03.041>
- [3] Lin, W.-C., Tsai, C.-F., Hu, Y.-H., & Jhang, J.-S. (2017). Clustering-based undersampling in class-imbalanced data. *Information Sciences*, 409–410, 17–26. <https://doi.org/10.1016/j.ins.2017.05.008>
- [4] Mortaz, E. (2020). Imbalance accuracy metric for model selection in multi-class imbalance classification problems. *Knowledge-Based Systems*, 210, 106490. <https://doi.org/10.1016/j.knsys.2020.106490>
- [5] Soltanzadeh, P., & Hashemzadeh, M. (2021). RCSMOTE: Range-Controlled synthetic minority oversampling technique for handling the class imbalance problem. *Information Sciences*, 542, 92–111. <https://doi.org/10.1016/j.ins.2020.07.014>
- [6] Wang, Y.-C., & Cheng, C.-H. (2021). A multiple combined method for rebalancing medical data with class imbalances. *Computers in Biology and Medicine*, 134, 104527. <https://doi.org/10.1016/j.combiomed.2021.104527>
- [7] Yuan, X., Xie, L., & Abouelenien, M. (2018). A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. *Pattern Recognition*, 77, 160–172. <https://doi.org/10.1016/j.patcog.2017.12.017>

● **11% Overall Similarity**

Top sources found in the following databases:

- 10% Publications database
- Crossref database
- Crossref Posted Content database
- 8% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Kelvin Leonardi Kohsasih, Zakarias Situmorang. "Analisis Perbandinga... Crossref	2%
2	Admi Syarif, Arafia Isnayu Akaf, Rizky Prabowo, Kurnia Muludi. "Analisi... Crossref	2%
3	Hartono Hartono, Opim Salim Sitompul, Tulus Tulus, Erna Budhiarti Na... Crossref	2%
4	Universitas Pamulang on 2023-03-28 Submitted works	1%
5	Sahat Saragih, Vira Afriati. "Peningkatan Pemahaman Konsep Grafik F... Crossref	1%
6	Alfa Saleh, Fina Nasari. "PENERAPAN EQUAL-WIDTH INTERVAL DISCR... Crossref	1%
7	York University on 2024-02-02 Submitted works	1%
8	Hartono, Erianto Ongko, Opim Salim Sitompul, Tulus, Erna Budhiarti N... Crossref	<1%