

Tersedia online di www.journal.unipdu.ac.id
UnipduHalaman jurnal di www.journal.unipdu.ac.id/index.php/register

Penanganan *imbalance class data* laboratorium kesehatan dengan *Majority Weighted Minority Oversampling Technique*

Meida Cahyo Untoro^a, Joko Lianto Buliali^b^{a,b} Teknik Informatika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesiaemail: ^a meidacahyountoro@gmail.com, ^b joko@cs.its.ac.id

INFO ARTIKEL

Sejarah artikel:Menerima 31 Mei 2018
Revisi 1 Agustus 2018
Diterima 1 Agustus 2018
Online 7 Agustus 2018**Kata kunci:**data laboratorium kesehatan
imbalanced
klasifikasi
MWMOTE**Keywords:***classification*
data laboratory health
imbalanced
MWMOTE**Style APA dalam mensitasi artikel ini:**Untoro, M. C. (2018). Penanganan *imbalance class data* laboratorium kesehatan dengan *Majority Weighted Minority Oversampling Technique*. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 4(1), 23-29.

ABSTRAK

Diagnosis suatu penyakit akan menjadi tepat jika didukung dengan berbagai proses mulai pengecekan awal (*amannesa*) sampai pengecekan laboratorium. Hasil dari proses laboratorium mempunyai informasi berbagai penyakit, akan tetapi beberapa jenis penyakit memiliki prevalensi rendah. Penyakit bervaleansi rendah memiliki pengaruh dalam penanganan pasien lebih lanjut. Dengan rasio yang tidak seimbang data laboratorium akan menyebabkan nilai akurasi menjadi rendah dalam pengklasifikasian dan penanganan penyakit. *Majority Weighted Minority Oversampling Technique* (MWMOTE) adalah salah satu cara untuk menyelesaikan *imbalanced*. Penelitian ini bertujuan menangani permasalahan ketidakseimbangan data laboratorium kesehatan sehingga diperoleh hasil pengklasifikasian penyakit dengan tingkat akurasi lebih tinggi. Hasil pada penelitian ini menunjukkan bahwa MWMOTE dapat meningkatkan akurasi untuk permasalahan ketidakseimbangan data sebesar 3,13%.

ABSTRACT

Diagnosis of a disease will be appropriate if supported by various processes ranging from initial checks (amannesa) to laboratory checks. Results from the laboratory process have information on various diseases, but some types of diseases have a low prevalence. Low-valvature disease has an effect in the treatment of the patient further. With an unbalanced ratio the laboratory data will cause the accuracy value to be low in the classification and handling of the disease. Majority Weighted Minority Oversampling Technique (MWMOTE) is one way to complete imbalanced. This study aims to address the problem of imbalance of health laboratory data to obtain the results of the classification of disease with a higher degree of accuracy. The results of this study indicate that MWMOTE can improve accuracy for data imbalance problems by 3.13%.

© 2018 Register: Jurnal Ilmiah Teknologi Sistem Informasi. Semua hak cipta dilindungi undang-undang.

1. Pendahuluan

Laboratorium kesehatan digunakan untuk menunjang pelaksanaan atau penyelenggaraan pelayanan kesehatan. Laboratorium tanpa pengembangan ilmu dan teknologi berdampak buruk terhadap penyakit. Data suatu penyakit dapat memberikan informasi untuk melakukan penyembuhan dan pemulihan terhadap pasien yang menderitanya. Peningkatan ilmu pengetahuan dan teknologi berimbang dapat memberikan kontribusi yang cukup baik dalam mempelajari pesanyakit-penyakit tertentu. Berbagai penyakit dapat dilihat karakteristiknya dari hasil laboratorium (leukosit, hemoglobin, dan glukosa). Diagnosis dapat ditegakkan berdasarkan hasil laboratorium pasien yang diperoleh dari pemeriksaan lebih lanjut. Tujuan dari pemeriksaan lanjut dapat mengetahui gejala penyakit yang sedang diderita misalnya penyakit diabetes, kanker payudara, dan leukimia (Meesad & Yen, 2003). *Data mining* mempermudah pengolahan data dalam jumlah besar, misalnya data laboratorium kesehatan, penarikan kesimpulan, dan acuan prediksi di masa mendatang (Kaur & Singh, 2014). Pada tahap *prepossessing data mining* membutuhkan ketelitian (Batra & Sachdev, 2016), yang selanjutnya akan dilakukan proses kalsifikasi dan *clustering* (Usharani & P.Sammulal, 2016). *Machine*

learning dan *data mining* memiliki kumpulan data yang terbagi menjadi satu atau lebih dari satu kelas, dan sering terjadi ketidakseimbangan kelas (*Imbalanced*) (Zhang, Krawczyk, Garcia, Rosales-Pérez, & Herrera, 2016). Tujuan dalam pengklasifikasian *dataset* memberikan prediksi terhadap data yang diujikan dan tidak mengenal *imbalanced* data. *Dataset* dianggap sebagai kelas mayoritas dan tidak memperhatikan kelas minoritas (Zheng, Cai, & Li, 2015).

Misclassification merupakan permasalahan yang sering terjadi pada pengklasifikasian data yang tidak seimbang dikarenakan *classifier* akan lebih condong terhadap data mayoritas sehingga didapat akurasi rendah pada data minoritas (Kim, Kang, and Kim 2015). Untuk menangani *imbalanced*, beberapa penelitian memanipulasi sampel data (pembuatan data sintetik) dan penggunaan algoritma (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). Metode klasifikasi memberikan nilai akurasi untuk seluruh data dengan meniadakan kelas minoritas dan seluruh data dianggap sebagai kelas mayoritas. *Dataset* diasumsikan memiliki distribusi yang *balance* serta kelas minoritas akan menjadi *noise* atau *outlier* (Napierała, 2012) (Phoungphol, 2013).

Oversampling merupakan salah satu cara untuk menangani permasalahan *imbalanced* dengan melakukan pendistribusian data yang seimbang dengan cara replikasi *instance* (data sintetik) minoritas secara acak dengan melakukan iterasi. *Oversampling* memiliki kekurangan dalam membuat data sintetik dengan munculnya *overfitting* karena mekanisme ini membuat data sintetik kurang tepat. *Majority Weighted Minority Oversampling Technique* (MWMOTE) dapat menangani *overfitting*. Pembuatan data sintetik pada MWMOTE terdapat 3 tahap, yaitu identifikasi sampel kelas minoritas dan kelas mayoritas pada *dataset*, pembobotan kelas minoritas, dan *clustering*. Hasil usulan tersebut ternyata mampu menurunkan derajat bias atau *noise*, serta menghasilkan data sintetik dengan tingkat akurasi lebih baik (Barua, Islam, Yao, & Murase, 2014).

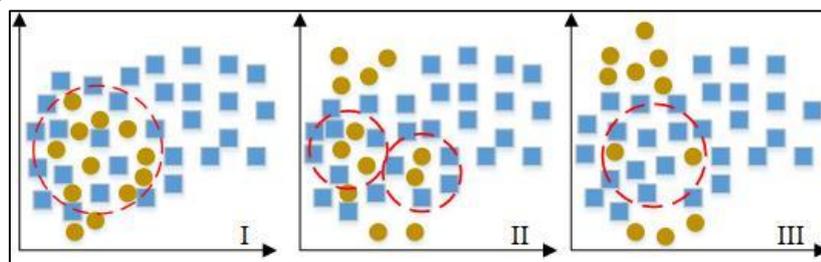
Penelitian ini bertujuan untuk menangani permasalahan *imbalanced* pada data laboratorium kesehatan dengan membuat data sintetik menggunakan MWMOTE yang representatif dengan melihat nilai *accuracy*, *precision*, *recall*, dan *f-measure*.

2. Penelitian Terkait

2.1. Imbalance

Data yang memiliki rasio yang tidak berimbang antara data satu dengan data lainnya dapat dikatakan sebagai *imbalanced*. *Data mining* mengartikan *imbalanced* dengan jumlah data kelas mayoritas lebih banyak dibandingkan dengan kelas minoritas. Permasalahan *imbalanced* terjadi pada *machine learning*, sehingga sering mengakibatkan *misclassification* berdampak pada nilai akurasi prediksi kelas menurun (Guo, Guo, Chen, & Jiang, 2016). Penurunan akurasi pada *imbalanced* disebabkan banyak ditemukan *noise* atau *outlier* pada *dataset* uji yang berasal dari kelas minoritas (Almeida, Barbosa, Pais, & Formosinho, 2007). Salah satu cara untuk menangani *imbalanced* dengan cara membandingkan metode klasifikasi dengan penambahan algoritma atau pemodifikasian metode (Seiffert, Khoshgoftar, & Hulse, 2009). *Imbalanced* dapat diselesaikan dengan penambahan data sintetik pada kelas minoritas dengan metode *oversampling* dan *undersampling*. *Imbalanced* memiliki kompleksitas yang cukup tinggi dan dibedakan ke dalam 3 kasus (Gambar 1) (Mahmood, 2015).

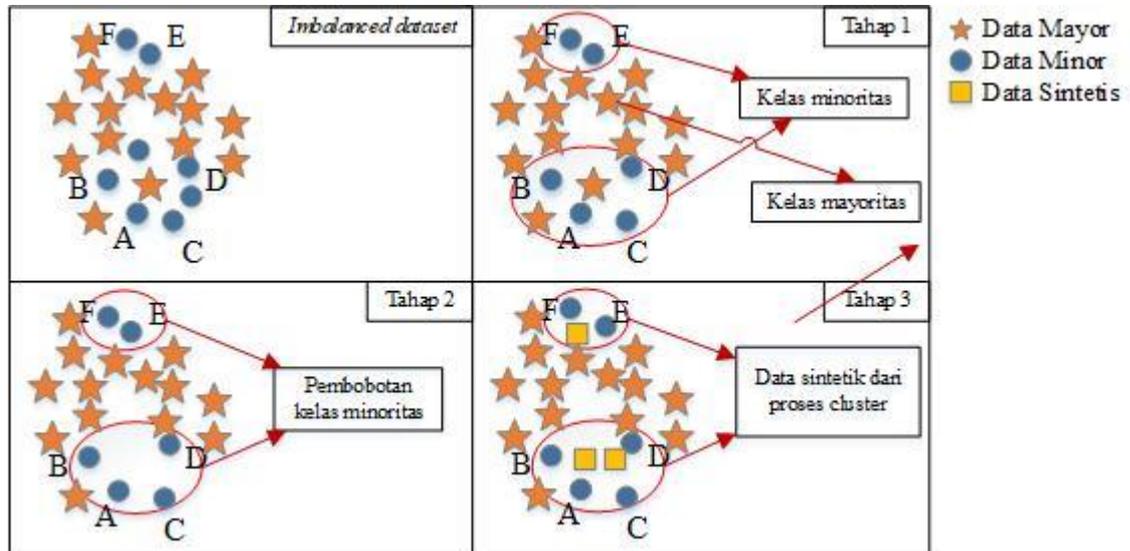
2.2. MWMOTE



Gambar 1. *Overlap* (I), *Small disjunct* (II), *Outlier/noise* (III)

Metode *oversampling* untuk pembuatan data sintetik mungkin memiliki beberapa ketidaktepatan dalam banyak skenario. Untuk mengatasi masalah ini, metode baru *Majority Weighted Minority Oversampling Technique* (MWMOTE) (Barua, Islam, Yao, & Murase, 2014). Tujuan dari MWMOTE ada dua, yaitu: Untuk meningkatkan proses pemilihan sampel dan untuk meningkatkan proses pembuatan sampel

sintetik. MWMOTE mempunyai 3 tahap (Gambar 2) yaitu: MWMOTE mengidentifikasi data minoritas yang sulit dipelajari. Data minoritas yang berada di dalam data mayoritas, data minoritas berdekatan (*borderline*) dengan data mayoritas serta data minoritas yang informasi berada pada *borderline*. Kedua, setiap anggota sampel informatif minoritas diberi bobot seleksi *sample weight* (Sw), sesuai dengan kepentingan data. Tahap terakhir adalah proses *clustering* bertujuan sampel data sintetik yang dihasilkan berada dalam satu kelompok *cluster*.



Gambar 2. Tahap MWMOTE

2.3. Laboratorium kesehatan

Laboratorium berasal dari bahasa latin yang berarti “tempat bekerja”. Laboratorium kesehatan melaksanakan pelayanan pemeriksaan di bidang Hematologi, Kimia Klinik, Mikrobiologi Klinik, Parasitologi Klinik, Imunologi Klinik atau bidang lain yang berkaitan dengan kepentingan kesehatan, dan bertujuan untuk menunjang diagnosis penyakit, penyembuhan penyakit dan pemulihan kesehatan (Usharani & P.Sammulal, 2016). Laboratorium klinik dan kesehatan memiliki banyak data untuk dijadikan informasi. Data yang didapat dari laboratorium akan disajikan berdasarkan kebutuhan masing-masing instansi (Gambar 3). Data yang akurat dan bagus akan memberikan informasi lebih tepat untuk mengambil sebuah keputusan terhadap pasien.



Gambar 3. Sysmax adalah alat untuk pengambilan data kesehatan

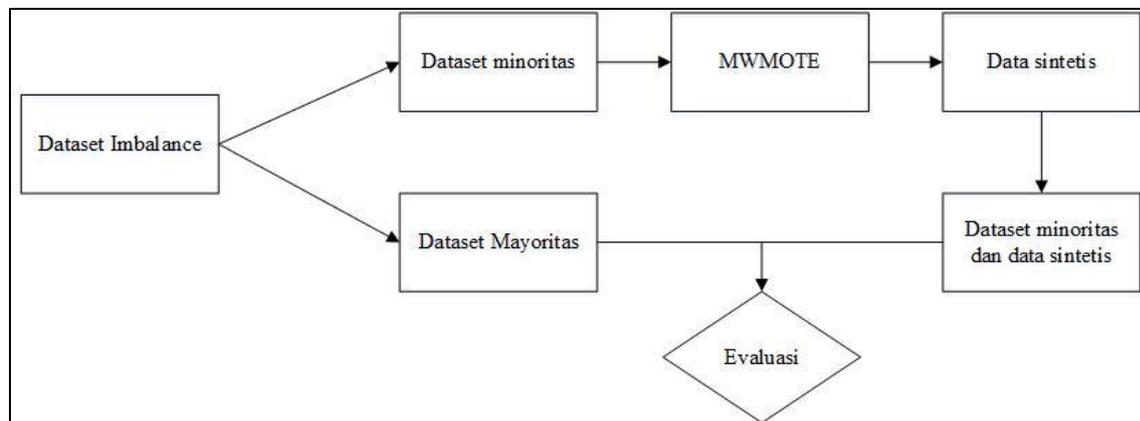
Data laboratorium yang digunakan sebagai rujukan bagi tenaga medis ahli dalam mendiagnosis penyakit yang diderita pasien. Tenaga medis masih banyak memiliki keterbatasan dalam mengambil keputusan terhadap diagnosis penyakit. Misalnya penyakit HIV atau kanker merupakan penyakit yang sangat jarang dijumpai dalam laboratorium. Diperlukan usulan penelitian untuk menangani permasalahan terkait jarang data kesehatan yang diderita oleh masyarakat. *Dataset* yang digunakan dalam penelitian ini adalah *dataset* laboratorium kesehatan di salah satu rumah sakit swasta yang memiliki karakteristik *imbalanced*. Data laboratorium memiliki rasio 2:8, di mana 2 merupakan penyakit yang jarang atau kelas minoritas dan 8 adalah kelas mayoritas. Jumlah *dataset* yang digunakan pada

penelitian sebanyak 388 baris data pada bulan November 2017. *Dataset* dibagi menjadi 2 bagian, data *training* sebesar 70% dan 30% untuk data *testing*. Tabel 1 memberikan informasi terkait atribut yang pada data laboratorium yang akan diteliti.

Tabel 1. Atribut *dataset* laboratorium kesehatan

No	Atribut	Definisi atribut
1	Hemoglobin	Media transpor oksigen dari paru paru ke seluruh jaringan tubuh.
2	Leukosit (<i>White Blood Cell</i> , WBC)	Berperanan dalam memerangi infeksi.
3	Trombosit (<i>Platelet</i>)	Membantu dalam proses pembekuan darah dan menjaga integritas <i>vascular</i> .
4	Eritrosit (<i>Red Blood Cell</i> , RBC)	Pengangkut/pembawa oksigen dari paru-paru untuk diedarkan ke seluruh tubuh.
5	Indeks Eritrosit (MCV, MCH, MCHC)	Membantu mendiagnosis penyebab anemia.
6	<i>Platelet Distribution Width</i> (PDW)	Koefisien variasi ukuran trombosit.
7	<i>Red Cell Distribution Width</i> (RDW)	Koefisien variasi dari volume eritrosit.

3. Metode Penelitian

Gambar 4. Diagram *block* penelitianTabel 2. *Dataset* laboratorium kesehatan

No	HGB	WBC	PLT	RBC	MCH	PDW	Class
1	15.3 g/dL	7.2 10 ³ /mm ³	358 10 ³ /mm ³	4.98 10 ⁶ /mm ³	30.6 Pg	16.30%	Positif
2	13.7 g/dL	10.1 10 ³ /mm ³	262 10 ³ /mm ³	4.31 10 ⁶ /mm ³	31.8 Pg	14.00%	Negatif
3	19.3 g/dL	9.1 10 ³ /mm ³	345 10 ³ /mm ³	6.41 10 ⁶ /mm ³	30.1 Pg	16.80%	Positif
4	15.3 g/dL	11.6 10 ³ /mm ³	297 10 ³ /mm ³	5.22 10 ⁶ /mm ³	29.3 Pg	14.50%	Positif
5	10.1 g/dL	14.4 10 ³ /mm ³	634 10 ³ /mm ³	3.56 10 ⁶ /mm ³	28.3 Pg	11.30%	Negatif

Gambar 4 merupakan usulan penelitian untuk menangani *imbalanced* data pada data laboratorium kesehatan. MWMOTE sebagai metode usulan yang digunakan untuk menangani permasalahan tersebut. Data laboratorium kesehatan bersifat *imbalanced* data, sebab dalam data laboratorium memiliki rasio penyakit yang tidak seimbang. Data yang dimiliki laboratorium mayoritas merupakan penyakit yang positif atau benar terserang penyakit, misalnya demam (*dengue*), diabetes, kolesterol dan lainnya. Sedangkan data penyakit yang masih dalam kondisi negatif atau perlu pemeriksaan lanjut (perlu dokter spesialis untuk menyatakan diagnosis itu benar positif). Salah satu penyakit hepatitis sangatlah jarang ditemui pada laboratorium kesehatan dan untuk hasil yang positif sangat jarang ditemui. Permasalahan yang dialami oleh laboratorium diperlukan penanganan terhadap *imbalance* data kesehatan. Metode MWMOTE digunakan untuk menyelesaikan permasalahan *imbalance* data. Dengan melakukan pemisahan data minoritas (*class* negatif) dengan data mayoritas (*class* positif) (Tabel 2). Data minoritas akan dijadikan model untuk pembuatan data sintetik dengan MWMOTE (Tabel 3). Data akan mendapatkan tambahan data sintetik untuk mendapatkan rasio yang sama antara data minoritas dengan data mayoritas. *Dataset* laboratorium memiliki rasio yang sama (*balance*) akan

dievaluasi menggunakan *classifier decision tree* (J48) dengan menggunakan aplikasi WEKA 3.6 dan didapat hasil evaluasi berdasarkan *accuracy*, *precision*, *recall*, dan *F-Measure*.

Tabel 3. Algoritma MWMOTE

1.	<i>S-maj</i> = Sampel kelas mayoritas
2.	<i>S-min</i> = Sampel kelas minoritas
3.	<i>N</i> = Sampel sintetik
4.	<i>k1</i> = Memprediksi <i>noise</i> pada kelas minoritas
5.	<i>k2</i> = Ketetanggaan kelas mayor yang digunakan sebagai batas kelas minor
6.	<i>k3</i> = Ketetanggaan kelas minor yang digunakan untuk membuat data sintetik

Fase 1.

1. *S-min* merupakan pencarian ketetanggaan yang terdekat berdasarkan jarak *Euclidean*.
2. *S-minf* merupakan pembuat kelas minor dari data sampel dan menghapus sampel kelas mayor.
3. *k2* untuk mencari ketetanggaan terdekat pada kelas mayor sesuai dengan jarak *euclidean*.
4. *S-bmaj* untuk menentukan *borderline* (batas) kelas mayor.
5. *k3* menentukan kelas minor yang akan digunakan untuk membuat data sintetik sesuai dengan jarak *Euclidean*.
6. *S-imin* menentukan kumpulan kelas minor yang informatif.

Fase 2

1. *I-w* untuk menentukan pembobototan pada anggota *S-imin*.
2. *S-w* untuk menyeleksi anggota dari kelas *I-w* untuk dijadikan kandidat sintetik.
3. *S-p* mengubah anggota *S-w* menjadi probabilitas.

Fase 3

1. *Lm* mencari *cluster* dari sampel *S-min*.
2. *S-omin* = *S-min* untuk membuat inisial.
3. Melakukan perulangan untuk memilih sampel dari *S-imin* sesuai distribusi probabilitas (*S-p*) misalnya anggota *cluster L-k*. Pilih sampel yang lain, secara acak dari anggota *cluster L-k*, dan menghasilkan data sintetik.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

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

$$F - \text{Measure} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Evaluasi dalam penanganan *imbalanced dataset* digunakan untuk mengetahui bagaimana metode usulan dapat menyelesaikan permasalahan tersebut. Evaluasi kinerja berdasarkan *precision*, *recall*, *F-Measure* dan *accuracy*. *True positive* (TP) adalah kasus di mana prediksi ya (memiliki penyakit), dan faktanya memang memiliki penyakit. *True negative* (TN) memperkirakan tidak, dan faktanya memang tidak memiliki penyakit. *False positive* (FP) memperkirakan ya, tetapi sebenarnya tidak memiliki penyakit (juga dikenal sebagai "Tipe I kesalahan"). *False Negative* (FN) memperkirakan tidak, tetapi faktanya adalah memiliki penyakit (juga dikenal sebagai "Kesalahan Tipe II"). *Accuracy* didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual (Persamaan 1). *Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi (Persamaan 2). *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem (Persamaan 3). *F-Measure* merupakan salah satu perhitungan evaluasi dalam temu kembali informasi yang mengkombinasikan *recall* dan *precision* (Persamaan 4).

4. Hasil dan Pembahasan

Data *imbalanced* terbagi mejadi data minoritas dan mayoritas. Data minoritas akan dilakukan proses *oversampling* untuk mendapatkan data sintetik. Di mana data *training* akan membuat pola untuk

pembentukan data sintetik dengan cara *oversampling* dan dievaluasi pada data *testing* untuk mengetahui kinerja dari metode yang diusulkan. Data *testing* merupakan data bagian dari data laboratorium secara keseluruhan tanpa ada penangan *imbalanced* dengan tujuan untuk mengevaluasi performa dari MWMOTE. Data sintetik yang terbentuk dari data minoritas sejumlah selisih antara data mayoritas dengan data minoritas. Penelitian tentang data laboratorium kesehatan akan memberikan suatu kontribusi penting bagi tenaga medis dan dunia kesehatan. Data laboratorium yang sangat beraneka ragam dan minim mengakibatkan tenaga medis menjadi susah dalam menggolongkan atau mengelompokkan pasien. Penyakit yang diderita pasien sangat bervariasi dan dalam penyakit tertentu hanya berjumlah sedikit berkisar 1% - 2% dari keseluruhan data penyakit. Tenaga medis di bagian laboratorium harus teliti dalam melakukan pemeriksaan untuk seluruh penyakit yang diderita pasien. Penyakit dapat dideteksi dari hasil laboratorium yang memiliki nilai dalam rentang tertentu untuk dinyatakan bahwa penyakit tersebut *positive* atau *negative*.

Tabel 4. Hasil *oversampling* dengan MWMOTE

	Tanpa MWMOTE				MWMOTE			
	Precision (-)	Recall (-)	F-Measure (-)	Accuracy (-)	Precision (-)	Recall (-)	F-Measure (-)	Accuracy(-)
P1	40.70%	57.90%	47.80%	81.25%	50.00%	68.40%	57.80%	85.16%
P2	43.50%	52.60%	47.60%	82.81%	52.00%	68.40%	59.10%	85.94%
P3	44.00%	57.90%	50.00%	82.81%	50.00%	68.40%	57.80%	85.16%
P4	40.60%	68.40%	51.00%	80.47%	53.80%	73.70%	62.20%	86.72%
P5	47.80%	57.90%	52.40%	84.38%	48.30%	73.70%	58.30%	84.38%
Avg	43.32%	58.94%	49.76%	82.34%	50.82%	70.52%	59.04%	85.47%

Pembuatan data sintetik pada metode MWMOTE mempunyai 3 tahap yaitu, proses pemisahan data minoritas dengan data mayoritas berdasarkan *class* yang sudah ditentukan pada tahap *preproses*. Data minoritas yang terpisah dari data mayoritas diberikan bobot, selanjutnya dilakukan *clustering* untuk mendapatkan data sintetik yang tidak menjadi *noise* atau *outlier*. Data sintetik akan berada dalam satu *cluster* data minoritas sehingga mempunyai karakteristik yang sama dengan kelas minoritas atau lebih representatif dan dapat meningkatkan hasil klasifikasi dengan menggunakan *decision tree* (J48). Hasil evaluasi *decision tree* pada proses *oversampling* data minoritas dengan menggunakan MWMOTE, dengan pembuatan data sintetik dari proses *oversampling* dilakukan 5 iterasi perulangan untuk mendapatkan hasil yang lebih representatif. Tabel 4 merupakan hasil evaluasi dari data *imbalanced* dengan dan tanpa menggunakan MWMOTE data laboratorium kesehatan. MWMOTE dapat memperbaiki permasalahan *imbalanced dataset* pada data laboratorium dengan nilai akurasi 85,47%, *precision* 50,82%, *recall* 70,52, dan *F-Measure* 59,04%.

5. Kesimpulan

Metode MWMOTE dapat menangani permasalahan ketidakseimbangan data dengan cara membuat data sintetik yang lebih representatif pada data laboratorium kesehatan. MWMOTE dapat dijadikan salah satu referensi dalam penanganan data *imbalance* pada permasalahan *data mining* dan *machine learning*. Hal tersebut membuktikan bahwa MWMOTE efektif meningkatkan performa klasifikasi data tidak seimbang.

6. Referensi

- Almeida, J., Barbosa, L., Pais, A., & Formosinho, S. (2007). Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. *Chemometrics and Intelligent Laboratory Systems*, 2007(2007), 208-217.
- Barua, S., Islam, M. M., Yao, X., & Murase, K. (2014). MWMOTE--Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning. *IEEE Transactions on Knowledge and Data Engineering*, 26(2), 405-425.
- Batra, S., & Sachdev, S. (2016). Organizing standardized electronic healthcare records data for mining. *Health Policy and Technology*, 5(3), 226-242.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *JAIR-Journal of Artificial Intelligence Research*, 16, 321-357.

- Fahrudin, T., Buliali, J. L., & Fatichah, C. (2016). Predictive modeling of the first year evaluation based on demographics data: Case study students of Telkom University, Indonesia. *2016 International Conference on Data and Software Engineering (ICoDSE)*. Denpasar: IEEE.
- Guo, S., Guo, D., Chen, L., & Jiang, Q. (2016). A centroid-based gene selection method for microarray data classification. *Journal of Theoretical Biology*, *400*(2016), 32-41.
- Kaur, B., & Singh, W. (2014). Review on Heart Disease Prediction System using Data Mining Techniques. *International Journal on Recent and Innovation Trends in Computing and Communication*, *2*(10), 3003 – 3008.
- Mahmood, A. M. (2015). Class Imbalance Learning in Data Mining – A Survey. *International Journal of Communication Technology for Social Networking Services*, *3*(2), 17-38.
- Meesad, P., & Yen, G. (2003). Combined numerical and linguistic knowledge representation and its application to medical diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *33*(2), 206-222.
- Napierała, K. (2012). *Improving Rule Classifiers For Imbalanced Data*. Poznań: Institute of Computing Science.
- Ng, W. W., Hu, J., Yeung, D. S., Yin, S., & Roli, F. (2015). Diversified Sensitivity-Based Undersampling for Imbalance Classification Problems. *IEEE Transactions on Cybernetics*, *45*(11), 2402-2412.
- Phoungphol, P. (2013). *A Classification Framework for Imbalanced Data*. Atlanta: Georgia State University. Retrieved from http://scholarworks.gsu.edu/cs_diss/78
- Seiffert, C., Khoshgoftaar, T. M., & Hulse, J. V. (2009). Hybrid sampling for imbalanced data. *Integrated Computer-Aided Engineering*, *16*(3), 193-210.
- Usharani, Y., & P.Sammulal. (2016). An Innovative Imputation and Classification Approach for Accurate Disease Prediction. *International Journal of Computer Science and Information Security (IJCSIS)*, *14*, 23-31.
- Zhang, Z., Krawczyk, B., Garcia, S., Rosales-Pérez, A., & Herrera, F. (2016). Empowering one-vs-one decomposition with ensemble learning for multi-class imbalanced data. *Knowledge-Based Systems*, *106*(15 August 2016), 251-263.
- Zheng, Z., Cai, Y., & Li, Y. (2015). Oversampling method for imbalanced classification. *Computing and Informatics*, *34*(5), 1017-1037.