



Available online at : <http://bit.ly/InfoTekJar>

## InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



Machine Learning

# Kombinasi Metode *Sampling* pada Pengklasifikasian Data Tidak Seimbang Menggunakan Algoritma *Support Vector Machine* (SVM)

Annisa Fadhillah Pulungan, Desilia Selvida

Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Medan, Indonesia

### KEYWORDS

Klasifikasi, *Oversampling*, *Undersampling*, *Confusion Matrix*, *Support Vector Machine*

### CORRESPONDENCE

Phone: +62 852 0643 0134

E-mail: [annisafpulungan@usu.ac.id](mailto:annisafpulungan@usu.ac.id)

### A B S T R A C T

Kelas tidak seimbang telah menjadi masalah pada pembelajaran mesin dan *data mining* selama beberapa tahun ini. Kelas tidak seimbang adalah kelas yang memiliki data yang tidak terdistribusi secara merata pada setiap kelasnya dimana terdapat satu kelas yang memiliki jumlah data yang lebih banyak dibanding kelas lainnya. Rasio ketidakseimbangan ini akan menyebabkan bias dalam kumpulan data pelatihan yang akan mempengaruhi kinerja pada pembelajaran mesin. Dimana menyebabkan terjadinya kesalahan klasifikasi pada kelas minoritas sehingga kelas minoritas dinyatakan sebagai kelas mayoritas. penelitian ini mencoba untuk menggunakan metode *Oversampling*, *Undersampling* dan kombinasi *Oversampling-Undersampling* pada data tidak seimbang sebagai pra-pemrosesan data untuk selanjutnya dilakukan klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Pada penelitian ini diperoleh kesimpulan bahwa penerapan metode *Oversampling*, *Undersampling* dan kombinasi *Oversampling-Undersampling* memiliki kemampuan yang baik dalam menangani ketidakseimbangan kelas dataset *Credit Card Fraud* walaupun pada nilai AUC, *Oversampling* memiliki nilai yang lebih rendah dibanding metode *Undersampling* dan kombinasi *Over-Undersampling*.

### PENDAHULUAN

Kelas tidak seimbang telah menjadi masalah pada pembelajaran mesin dan *data mining* selama beberapa tahun ini. Kelas tidak seimbang adalah kelas yang memiliki data yang tidak terdistribusi secara merata pada setiap kelasnya dimana terdapat satu kelas yang memiliki jumlah data yang lebih banyak dibanding kelas lainnya[1].

Keadaan kelas yang memiliki jumlah data lebih banyak disebut dengan kelas mayoritas dan keadaan kelas yang memiliki jumlah data yang lebih sedikit disebut dengan kelas minoritas. Perbedaan yang terlalu signifikan antara kelas mayoritas dan kelas minoritas akan menyebabkan

rasio tidak seimbang yang disebut dengan *Imbalance Rasio* [2].

Rasio ketidakseimbangan ini akan menyebabkan bias dalam kumpulan data pelatihan yang akan mempengaruhi kinerja pada pembelajaran mesin. Dimana menyebabkan terjadinya kesalahan klasifikasi pada kelas minoritas sehingga kelas minoritas dinyatakan sebagai kelas mayoritas. Hal ini tentu akan mempengaruhi hasil klasifikasi terkhusus pada kasus-kasus yang membutuhkan hasil klasifikasi yang tepat di beberapa bidang seperti diagnosis suatu penyakit pada bidang Kesehatan, pendeteksian penipuan kartu kredit online pada bidang perbankan, dan lain-lain [2].

Dalam memecahkan masalah ini, terdapat beberapa pendekatan yang dapat dilakukan. Salah satunya adalah dengan melakukan modifikasi dataset yaitu dengan menduplikasi kelas minoritas sehingga jumlah kelas minoritas sama dengan jumlah kelas mayoritas. Proses ini disebut dengan *Oversampling*. Selain itu dapat juga dilakukan dengan mereduksi kelas mayoritas sesuai dengan jumlah kelas minoritas. Proses ini disebut dengan *Undersampling* [3].

Penelitian oleh Vuttipittayamongkol P et al [4] menyatakan bahwa Teknik *Undersampling* digunakan menghilangkan tumpang tindih pada nilai kelas minoritas. Teknik *Undersampling* pada pengklasifikasian data tidak seimbang ini memberikan peningkatan yang signifikan dalam proses klasifikasinya secara statistik. Namun *Undersampling* memiliki kelemahan yaitu banyaknya sampel data yang terhapus dimana terdapat kemungkinan adanya kehilangan informasi penting pada data yang terhapus [2].

Penelitian yang dilakukan oleh Wenhao et al [5] dengan judul “*An Improved Oversampling Algorithm Based on the Samples Selection Strategy for Classifying Imbalanced Data*” yang bertujuan untuk meningkatkan metode *Oversampling* terhadap pengklasifikasian data tidak seimbang. Hasil penelitian ini menunjukkan bahwa algoritma *Oversampling* yang ditingkatkan menunjukkan kinerja yang baik pada klasifikasi data tidak seimbang.

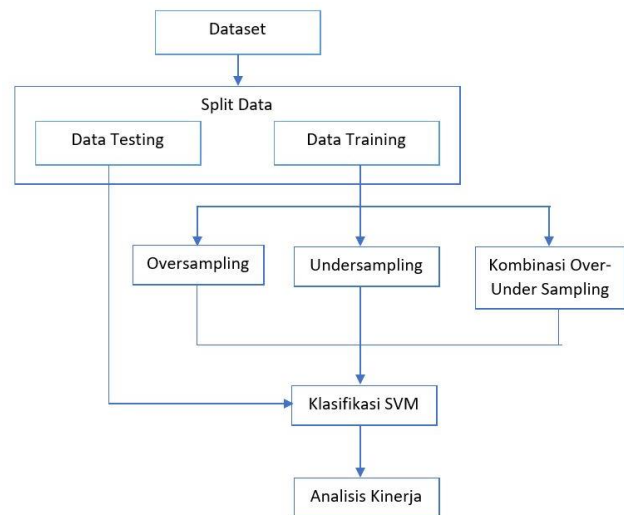
Rustam et al [6] juga melakukan penelitian terkait data tidak seimbang dimana penelitian ini mencoba melakukan *hybrid preprocessing* dengan menggabungkan metode *Oversampling* dan *Undersampling* pada proses pengklasifikasian pada penyakit Infark Serebral. Penelitian ini menghasilkan keakuratan metode yang lebih tinggi pada metode hibrida yaitu sebesar 94% dibandingkan menggunakan metode *Oversampling* saja atau *Undersampling* saja.

Berdasarkan beberapa penelitian terdahulu, penelitian ini mencoba untuk menggunakan metode *Oversampling*, *Undersampling* dan kombinasi *Over-Undersampling* pada data tidak seimbang sebagai pra-pemrosesan data untuk selanjutnya dilakukan klasifikasi menggunakan algoritma *Support Vector Machine (SVM)*. Tujuan utama penelitian ini adalah untuk menemukan metode yang tepat pada pra-pemrosesan data pada pengklasifikasian data tidak seimbang. Perhitungan Kinerja yang akan digunakan pada penelitian ini adalah *G-Means*, *F1 Score*, Luas AUC dan ROC.

## METODE

### Alur Penelitian

Langkah kerja yang dilakukan dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Alur Kerja Penelitian

### Dataset

Dataset pada penelitian ini menggunakan dataset *Credit Card Fraud* yang diambil dari *Kaggle*. Dataset ini berisi transaksi kartu kredit oleh pemegang kartu Eropa pada September 2013. Dataset ini terdiri dari 284.807 transaksi yang terdiri atas 2 kelas yaitu kelas 1 dengan 492 transaksi penipuan dan kelas 0 dengan 284.315 data transaksi benar.

### Data Splitting

Pada tahap *data splitting* dilakukan pembagian data ke dalam dua bagian yaitu data *testing* dan data *training*. Tabel 1 merupakan ilustrasi pembagian data pada proses *data splitting*.

Tabel 1. Ilustrasi *Data Splitting*

Dataset	
<i>Data Training</i> = 75 %	<i>Data Testing</i> = 25 %

Pada tahap ini, *data training* akan digunakan sebagai representasi pengetahuan pada proses pembelajaran oleh algoritma klasifikasi untuk memprediksi suatu kelas pada data baru yang belum dikenali. Pada penelitian ini, data training yang akan digunakan sebesar 75% dari jumlah dataset keseluruhan. Sedangkan *data testing* akan digunakan untuk mengukur sejauh mana model klasifikasi pada proses *training* dapat mengklasifikasikan suatu data dengan benar. *Data testing* yang akan digunakan sebesar 25% dengan jumlah data sebesar 71.082 untuk data kelas 0 dan 120 untuk data kelas 1 [2].

## Metode Sampling

### Klasifikasi Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran yang terawasi. Tujuan dari SVM adalah untuk meminimalisir resiko struktural dan memperhitungkan aspek general dala menemukan *hyperplane* terbaik untuk memisahkan suatu data dari kelas yang ditentukan. *Hyperplane* dikatakan baik apabila memiliki *margin* yang besar diikuti dengan *error* yang kecil. *Margin* adalah jarak antar *hyperplane* suatu kelas dengan *hyperplane* kelas lain. Titik data suatu kelas yang paling dengan dengan *hyperplane* merupakan kelas *hyperplane* yang disebut juga dengan *support vector* [6].

Persamaan *Lagrangian* merupakan bentuk dual dari SVM yang dapat memecahkan masalah SVM melalui *quadratic programming* [2].

$$X_n = \frac{0.8*(x-a)}{b-a} + 0.1 \quad (1)$$

Keterangan :

$X_n$  = nilai ke-n

A = nilai angka terendah

B = nilai angka tertinggi

0.8/0.1 = ketepatan

### Perhitungan Kinerja Algoritma

Perhitungan kinerja algoritma memiliki peran yang sangat penting dalam mengevaluasi kinerja suatu algoritma pada *machine learning*. Beberapa perhitungan kinerja yang sering digunakan dalam mengukur tingkat keberhasilan algoritma adalah akurasi. Namun, perhitungan akurasi pada pengklasifikasian data tidak seimbang tidak dapat dijadikan patokan utama berhasilnya algoritma dalam melakukan klasifikasi.

Hal ini tidak sesuai pada dataset yang tidak seimbang karena akurasi 99% dapat dicapai ketika rasio ketidakseimbangan sebesar 1:99 dimana algoritma klasifikasi mendiskriminasi semua kelas termasuk kelas minoritas sebagai kelas mayoritas [8]. Untuk itu dalam penelitian ini, penulis melakukan perhitungan kinerja berdasarkan *Confusion Matrix*, *G-Means*, *F1-Score*, *ROC* dan *AUC*.

### Confusion Matrix

Pada umumnya hasil kinerja algoritma pengklasifikasi akan dievaluasi menggunakan *confusion matrix*. Pada kasus pengklasifikasian kelas biner, maka matriks yang dibentuk adalah 2x2 dimana kolom menyatakan hasil

prediksi algoritma klasifikasi dan baris menyatakan nilai sebenarnya suatu kelas yang ditunjukkan pada Tabel 2 [9].

Tabel 2. *Confusion Matrix*

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Actual Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Berikut penjelasan terkait table confusion matrix adalah :

*True positive (TP)* : jumlah kasus positif yang diklasifikasi sebagai kelas positif

*False Negative (FN)* : jumlah kasus positif yang diklasifikasi sebagai kelas negative

*False Positive (FP)* : jumlah kasus negative yang diklasifikasikan sebagai kelas positif

*True Negative (TN)* : jumlah kasus negative yang diklasifikasikan sebagai kelas negatif

Beberapa kinerja evaluasi yang dapat dihasilkan dari *confusion matrix* antara lain [9]:

Akurasi (*Accuracy*) adalah matriks evaluasi yang digunakan untuk mewakili nilai keakuratan model dalam mengklasifikasikan dengan benar.

$$\text{Akurasi} = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (2)$$

Presisi (*Precision*) adalah perbandingan prediksi benar positif dengan keseluruhan data yang diprediksi positif.

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

Recall (*Sensitivity*) adalah perbandingan ukuran kebenaran positif yang diprediksi positif secara tepat.

$$\text{Recall} = \frac{(TP)}{(TP+FN)} \quad (4)$$

*Specificity* adalah perbandingan prediksi benar negatif dengan keseluruhan data negatif.

$$\text{Specificity} = \frac{(TN)}{(TN+FP)} \quad (5)$$

### G-Means

*G-Means* merupakan parameter evaluasi yang digunakan untuk mengukur kinerja klasifikasi pada dua kelas yaitu sensitifitas (akurasi pada kelas positif) dan spesifitas (akurasi pada kelas negatif)[9]. Matriks ini cocok digunakan untuk mengevaluasi masalah data tidak seimbang sebab matriks ini menunjukkan keseimbangan antara kinerja pada kelas mayoritas dan minoritas. Nilai *G-Mean* yang rendah menunjukkan bahwa kinerja algoritma klasifikasi yang rendah terhadap suatu kelas[8].

$$G - \text{Means} = \sqrt{\text{Sensitivity} * \text{Specificity}} \quad (6)$$

### F1-Score

F1-Score adalah hubungan perbandingan rata-rata presisi dan *recall*. Hasil klasifikasi terbaik akan memiliki nilai F1-Score mendekati 1 dan hasil klasifikasi terburuk mendekati nilai 0 [3].

$$f1 = 2 \times \left( \frac{\text{Presisi} * \text{Recall}}{\text{Presisi} + \text{Recall}} \right) \quad (7)$$

### ROC dan AUC

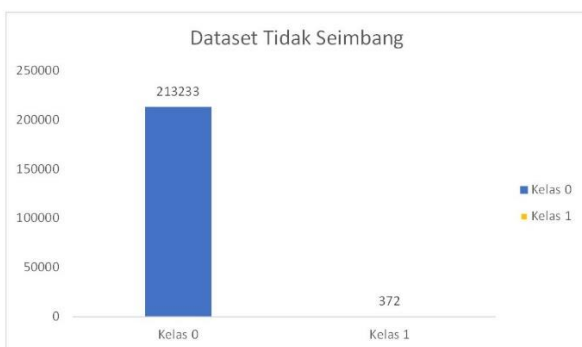
ROC (*Receiver Operating Characteristic*) Curve adalah kurva yang memberikan nilai *true positive rate* dan *false positive rate* pada kelas yang sama. Semakin condong arah kurva mendekati angka 1 maka semakin baik kemampuan pengklasifikasian suatu algoritma dalam membedakan antara kelas positif dan kelas negatif [9].

AUC (*Area Under Curve*) adalah luas area dibawah *curve* ROC dimana menjadi indikator ringkasan kinerja kurva ROC yang dapat menghitung kinerja klasifikasi pada satu matriks. AUC dapat dihitung dengan berbagai teknik, namun paling banyak digunakan adalah metode trapezium. Secara umum, variasi nilai AUC berada pada rentang nilai 0,5-1 [9].

## HASIL DAN DISKUSI

Penelitian ini menggunakan metode *Oversampling*, *Undersampling* dan kombinasi *Over-Undersampling* pada klasifikasi data tidak seimbang menggunakan algoritma *support vector machine* (SVM). Kemudian membandingkan untuk mendapatkan hasil kinerja yang baik pada data tidak seimbang. Perhitungan Kinerja yang akan digunakan pada penelitian ini adalah *G-Means*, *F1 Score*, Luas AUC dan ROC.

Dataset yang digunakan pada penelitian ini memiliki rasio ketidakseimbangan antar kelas yaitu sebesar 500. Gambar 2 menunjukkan jumlah data yang tidak seimbang antar kelas.



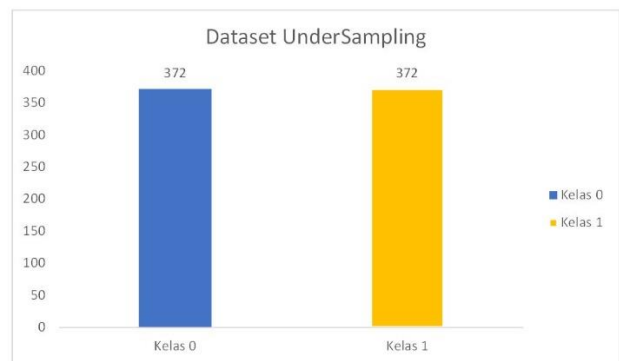
Gambar 2. Dataset tidak seimbang

Parameter yang digunakan pada proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) pada penelitian ini adalah

Tipe SVM	: C-Classification
SVM-Kernel	: radial
Cost	: 1

### Analisis Metode Undersampling

Pada proses *training*, jumlah data set yang digunakan sebesar 75% dari total dataset dimana terdapat 213.233 data untuk kelas 0 dan 372 data untuk kelas 1. Pada proses *Undersampling* dilakukan penghapusan sebagian dataset mayoritas sehingga didapatkan jumlah yang sama dengan kelas minoritas. Gambar 3 menunjukkan diagram keseimbangan antar kelas pada metode *Undersampling*.



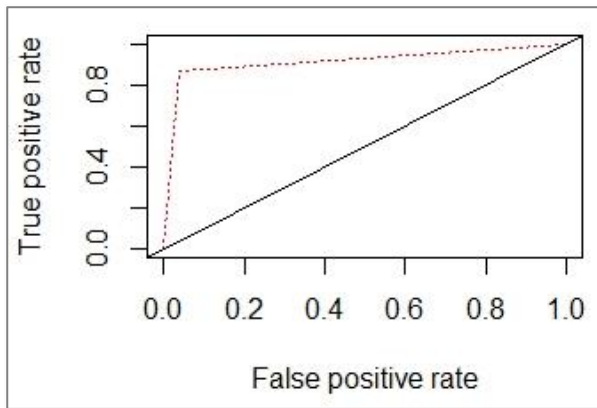
Gambar 3. Dataset *Undersampling*

Dari Gambar 3 dapat dilihat bahwa Sebagian besar dataset kelas mayoritas sudah dihapus sehingga kelas mayoritas dan kelas minoritas memiliki jumlah dataset yang sama yaitu sebesar 372. Tabel 3 menunjukkan hasil kinerja *confusion matrix* dari metode *Undersampling* pada data *testing*.

Tabel 3. Kinerja *Undersampling*

Kinerja	Hasil
<i>Sensitivity/Recall</i>	0.99
<i>Specificity</i>	0.038
<i>G-Means</i>	0.172
<i>Precision</i>	0.963
<i>F1-Score</i>	0.98

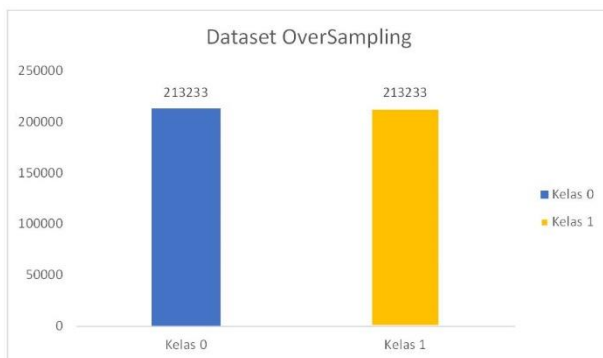
Gambar 4 menunjukkan gambar kurva *Receiver Operating Characteristic* (ROC) pada metode *Undersampling*.

Gambar 4. Diagram ROC metode *Undersampling*

Pada gambar 4 dapat dihitung luas AUC pada kurva ROC yaitu sebesar 0.912.

### Analisis Metode Oversampling

Pada proses *Oversampling* dilakukan duplikasi dataset minoritas sehingga didapatkan jumlah yang sama dengan kelas mayoritas. Gambar 5 menunjukkan diagram keseimbangan antar kelas pada metode *Oversampling*

Gambar 5. Dataset *Oversampling*

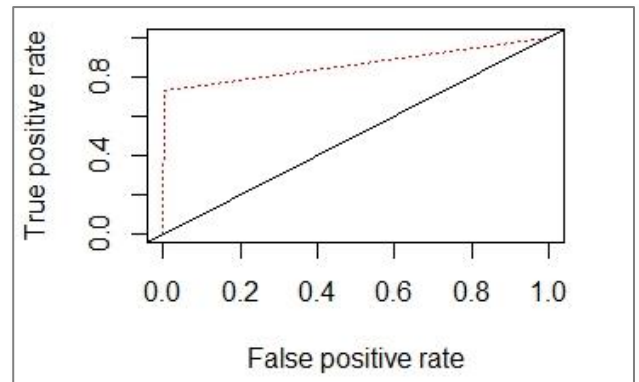
Dari Gambar 5 dapat dilihat bahwa Sebagian besar dataset kelas minoritas sudah diduplikasi sehingga kelas mayoritas dan kelas minoritas memiliki jumlah dataset yang sama yaitu sebesar 213.233. Tabel 4 menunjukkan hasil kinerja *confusion matrix* dari metode *Oversampling* pada data *testing*.

Tabel 4. Kinerja *Oversampling*

Kinerja	Hasil
<i>Sensitivity/Recall</i>	0.99
<i>Specificity</i>	0.31
<i>G-Means</i>	0.554
<i>Precision</i>	0.997
<i>F1-Score</i>	0.99

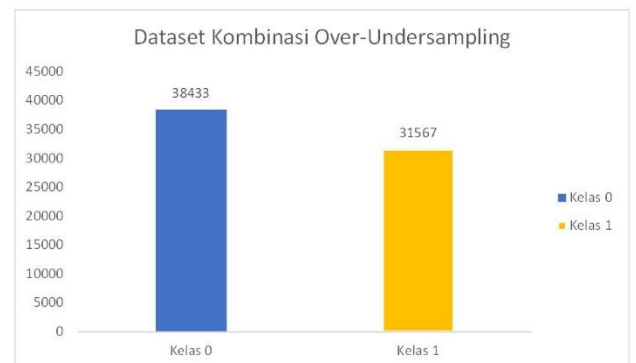
Gambar 6 menunjukkan gambar kurva *Receiver Operating Characteristic* (ROC) pada metode

*Oversampling* dengan luas AUC pada Kurva ROC sebesar 0.865.

Gambar 6. Diagram ROC metode *Oversampling*

### Analisis Metode Kombinasi Under-Oversampling

Pada proses kombinasi ini dilakukan penghapusan Sebagian dataset mayoritas dan penduplikasian dataset minoritas sehingga didapatkan jumlah yang sama dengan kelas minoritas. Gambar 7 menunjukkan diagram keseimbangan antar kelas pada metode kombinasi *Under-Oversampling*.

Gambar 7. Dataset Kombinasi *Under-Oversampling*

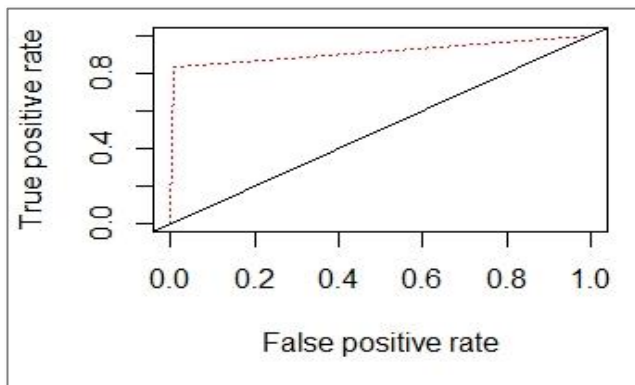
Dari Gambar 7 dapat dilihat bahwa Sebagian besar dataset kelas mayoritas sudah dihapus dan terdapat duplikasi dataset minoritas sehingga didapat jumlah dataset antar kelas tidak memiliki rasio ketidakseimbangan yang tinggi. Tabel 5 menunjukkan hasil kinerja *confusion matrix* dari metode kombinasi pada data *testing*.

Tabel 5. Kinerja Kombinasi

Kinerja	Hasil
<i>Sensitivity/Recall</i>	0.99
<i>Specificity</i>	0.27
<i>G-Means</i>	0.37
<i>Precision</i>	0.99
<i>F1-Score</i>	0.99



Gambar 8 menunjukkan gambar kurva *Receiver Operating Characteristic* (ROC) pada metode kombinasi dengan luas AUC pada kurva ROC sebesar 0.914.



Gambar 8. Diagram ROC metode kombinasi

Tabel 6 menunjukkan perbandingan hasil kinerja *confusion matriks* dari metode *Undersampling*, *Oversampling* dan kombinasi *Over-Undersampling*.

Tabel 6. Perbandingan Hasil Kinerja *Confusion Matrix*

Metode	G-Means	F1-Score	Luas AUC
<i>Undersampling</i>	0.172	0.98	0.912
<i>Oversampling</i>	0.554	0.99	0.865
Kombinasi	0.372	0.99	0.914

Dari tabel diatas dapat dilihat bahwa metode *Oversampling* memiliki nilai *G-Means* dan *F1-Score* yang lebih tinggi dibanding pada metode *Undersampling* dan Kombinasi yaitu sebesar 0.554 untuk nilai *G-Means* dan 0.99 untuk nilai *F1-Score*. Namun memiliki nilai AUC yang lebih rendah dibandingkan kedua metode lainnya yaitu sebesar 0.865. Selanjutnya diikuti dengan metode kombinasi yang memiliki nilai *G-Means* sebesar 0.372, *F1-Score* sebesar 0.99 dan nilai luas AUC sebesar 0.914. Dan metode *Undersampling* yang memiliki nilai *G-Means* sebesar 0.172, *F1-Score* sebesar 0.98 dan luas AUC sebesar 0.912.

## KESIMPULAN

Dataset *Credit Card Fraud* adalah dataset yang memiliki ketidakseimbangan kelas yang cukup tinggi yang memiliki rasio ketidakseimbangan sebesar 1:500. Metode *sampling* yang digunakan pada penelitian ini antara lain *Undersampling*, *Oversampling* dan kombinasi *Over-Undersampling* untuk menangani ketidakseimbangan kelas pada proses klasifikasi dataset menggunakan algoritma *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa metode *Oversampling* memiliki nilai *G-Means* dan *F1-Score* yang lebih tinggi dibanding dengan kedua metode lainnya yaitu sebesar

0.554 untuk nilai *G-Means* dan 0.99 untuk nilai *F1-Score*. Namun memiliki nilai AUC yang lebih rendah dibandingkan kedua metode lainnya yaitu sebesar 0.865 untuk nilai AUC *Oversampling*, 0.914 untuk nilai AUC kombinasi dan 0.912 untuk nilai AUC *Undersampling*.

Berdasarkan hasil eksperimen diatas diperoleh kesimpulan bahwa penerapan metode *Oversampling*, *Undersampling* dan kombinasi *Over-Undersampling* memiliki kemampuan yang baik dalam menangani ketidakseimbangan kelas dataset *Credit Card Fraud* walaupun pada nilai AUC, *Oversampling* memiliki nilai yang lebih rendah dibanding metode *Undersampling* dan kombinasi *Over-Undersampling*.

## REFERENSI

- [1] Siringoringo, R. (2018). Jurnal ISD : Klasifikasi Data tidak Seimbang Menggunakan Algoritma SMOTE dan KNN, 3(1), 44-49.
- [2] Gagah Gumelar, Norlaila2, Quratul Ain, Riza Marsuciati, Silvi Agustanti Bambang, Andi Sunyoto, & M. Syukri Mustafa. (2021). Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance. *Prosiding SISFOTEK*, 5(1), 250 - 255.
- [3] Rustam, Zuherman & Utami, Dea & Hidayat, Rahmat & Pandelaki, Jacob & Nugroho, Widy. (2019). Hybrid Preprocessing Method for Support Vector Machine for Classification of Imbalanced Cerebral Infarction Datasets. *International Journal on Advanced Science, Engineering and Information Technology*. 9. 685.
- [4] Vuttipittayamongkol P., Elyan E., Petrovski A., Jayne C. (2018) Overlap-Based Sampling for Improving Imbalanced Data Classification. In: Yin H., Camacho D., Novais P., Tallón-Ballesteros A. (eds) *Intelligent Data Engineering and Automated Learning – IDEAL 2018*. IDEAL 2018. Lecture Notes in Computer Science, vol 11314. Springer, Cham. [https://doi.org/10.1007/978-3-030-03493-1\\_72](https://doi.org/10.1007/978-3-030-03493-1_72).
- [5] Wenhao Xie, Gongqian Liang, Zhonghui Dong, Baoyu Tan, Baosheng Zhang, "An Improved Oversampling Algorithm Based on the Samples' Selection Strategy for Classifying Imbalanced Data", *Mathematical Problems in Engineering*, vol. 2019, ArticleID 3526539, 13 pages, 2019 . <https://doi.org/10.1155/2019/3526539>
- [6] Rustam, Z., Utami, D. A., Hidayat, R., Pandelaki, J., & Nugroho, W. A. (2019). Hybrid preprocessing method for support vector machine for classification of imbalanced cerebral infarction datasets. *International Journal on Advanced Science Engineering Information Technology*, 9(2).
- [7] Syukron, Akhmad & Subekti, Agus. (2018). Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit. *Jurnal Informatika*. 5. 175-185. 10.31311/ji.v5i2.4158.
- [8] Fujiwara K, Huang Y, Hori K, Nishioji K, Kobayashi M, Kamaguchi M, Kano M. Over- and Under-sampling Approach for Extremely Imbalanced and Small Minority Data Problem in Health Record Analysis. *Front Public Health*. 2020 May 19;8:178. doi:

- [9] 10.3389/fpubh.2020.00178. PMID: 32509717; PMCID: PMC7248318.
- [10] Bekkar, Mohamed & Djema, Hassiba & Alitouche, T.A.. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*. 3. 27-38.