

## Penanganan Ketidak Seimbangan Kelas Menggunakan Pendekatan Level Data

Abdurraghib Segaf Suweleh, Dyah Susilowati, Hairani, Khairan Marzuki

Universitas Bumigora, Indonesia

### Article Info

#### Article history:

Received, 14 August 2020

Revised, 21 August 2020

Accepted, 11 September 2020

#### Kata kunci:

Ketidak Seimbangan Kelas

Level Data

SMOTE

K-Means-SMOTE

Metode C4.5

### ABSTRAK

Setiap tahun bagian kemahasiswaan Universitas Bumigora melakukan seleksi mahasiswa yang berhak mendapatkan Beasiswa Peningkatan Prestasi Akademik (Beasiswa PPA). Dalam proses seleksi pemilihan penerima Beasiswa PPA terdapat permasalahan seperti kesulitan dalam menentukan mahasiswa yang berhak menerima beasiswa, dikarenakan jumlah kuota beasiswa lebih sedikit dibandingkan jumlah mahasiswa yang mendaftar beasiswa. Kumpulan data hasil seleksi Beasiswa PPA sebanyak 150 *instance*. Terdapat ketidak seimbangan kelas pada data yang digunakan yaitu 85 *instance* kelas tidak layak dan 65 *instance* kelas layak. Solusi yang ditawarkan adalah menggunakan pendekatan level data untuk menyeimbangkan kelasnya seperti metode SMOTE dan k-means-SMOTE. Adapun tujuan penelitian ini adalah menangani permasalahan ketidak seimbangan kelas pada data beasiswa PPA Universitas Bumigora menggunakan pendekatan level data untuk meningkatkan kinerja metode C4.5. Tahapan-tahapan penelitian ini terdiri dari pengumpulan data Beasiswa PPA, data *preprocessing*, klasifikasi, dan pengujian kinerja. Berdasarkan hasil pengujianya, pendekatan level data menggunakan metode k-means-SMOTE dan metode C4.5 memiliki kinerja terbaik untuk klasifikasi penerima Beasiswa PPA dengan akurasi 81.3%, sensitivitas 84.9%, dan spesifisitas 77.6%. Dengan demikian, metode k-means-SMOTE dan metode C4.5 memiliki kinerja terbaik berdasarkan akurasi, sensitivitas, dan spesifisitas.

#### Keywords:

Class Imbalance

Data level

SMOTE

K-Means-SMOTE

C4.5 Method

### ABSTRACT

Every year student affairs division of Bumigora University selects students who are entitled to receive PPA scholarships. In selection process for PPA scholarship recipients there are problems such as difficulties in determining which students are eligible to receive scholarships, because number of scholarship quotas is less than number of students applying for scholarships. The data collection resulted from PPA scholarship selection is 150 instances. There is a class imbalance in data used, namely 85 instances of majority class and 65 instances of minority class. Solution offered is to use a data level approach to balance the class such as SMOTE and k-means-SMOTE methods. The purpose of this study is to address problem of class imbalance in PPA Bumigora University scholarship data using a data level approach to improve performance of C4.5 method. The stages of this research consisted of collecting PPA scholarship data, preprocessing data, classification, and performance testing. Based on test results, data level approach using k-means-SMOTE method and C4.5 method has the best performance for classification of PPA scholarship recipients with an accuracy of 81.3%, sensitivity of 84.9%, and specificity of 77.6%. Thus, k-means-SMOTE method and C4.5 method have the best performance based on accuracy, sensitivity, and specificity.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



#### Penulis Korespondensi:

Hairani,

Program Studi Ilmu Komputer,

Universitas Bumigora.

Email: hairani@universitasbumigora.ac.id

## 1. PENDAHULUAN

Beasiswa Peningkatan Prestasi Akademik atau PPA merupakan beasiswa yang diberikan oleh pemerintah kepada mahasiswa yang memiliki prestasi dibidang akademik. Universitas Bumigora merupakan salah satu kampus swasta yang ikut berperan mendukung pemerintah dalam hal pemberian beasiswa PPA kepada mahasiswanya. Setiap tahun bagian kemahasiswaan melakukan seleksi mahasiswa yang berhak mendapatkan beasiswa PPA. Berdasarkan hasil wawancara dengan Wakil Rektor 3 Universitas Bumigora terdapat permasalahan dalam proses seleksi pemilihan penerima beasiswa PPA seperti kesulitan menentukan mahasiswa yang berhak menerima beasiswa, dikarenakan jumlah kuota beasiswa lebih sedikit dibandingkan jumlah mahasiswa yang mendaftar beasiswa. Kumpulan data hasil seleksi beasiswa PPA yang dilakukan setiap tahunnya, hanya tersimpan pada komputer dan belum dimanfaatkan dengan maksimal. Padahal kumpulan data tersebut dapat dianalisis untuk mendapatkan pola-pola atau pengetahuan seperti kriteria mahasiswa yang mendapatkan beasiswa PPA atau tidak. Salah satu solusi yang ditawarkan pada penelitian ini adalah menggunakan konsep *data mining*. *Data mining* adalah sebuah cara untuk mengubah kumpulan-kumpulan data yang ada menjadi sebuah pengetahuan sehingga mendapatkan pola-pola tertentu. Salah satu *task data mining* adalah klasifikasi. Seleksi penerima beasiswa sudah banyak dilakukan penelitian sebelumnya dengan menggunakan berbagai metode *data mining* seperti metode C4.5 [1]–[5], *k-Nearest Neighbours* (k-NN)[6]–[9], *Bayesian MCMC* [10], dan *Naive Bayes* [11].

Penelitian [3] menggunakan metode C4.5 untuk klasifikasi penerima beasiswa PPA. Kelemahan penelitian tersebut adalah tidak menyelesaikan permasalahan ketidakseimbangan kelas pada data beasiswa yang digunakan. Penelitian [11] untuk klasifikasi penerima beasiswa bidikmisi menggunakan metode *Naive Bayes*. Kelemahan penelitian tersebut, tidak menyelesaikan permasalahan ketidakseimbangan kelas pada data yang digunakan dan tidak hanya itu, penelitian tersebut menggunakan metode pengujian data *split* dengan menentukan jumlah *data training* dan *data testing*, hal tersebut akan berdampak pada hasil klasifikasi yang rawan terjadinya *overfitting*.

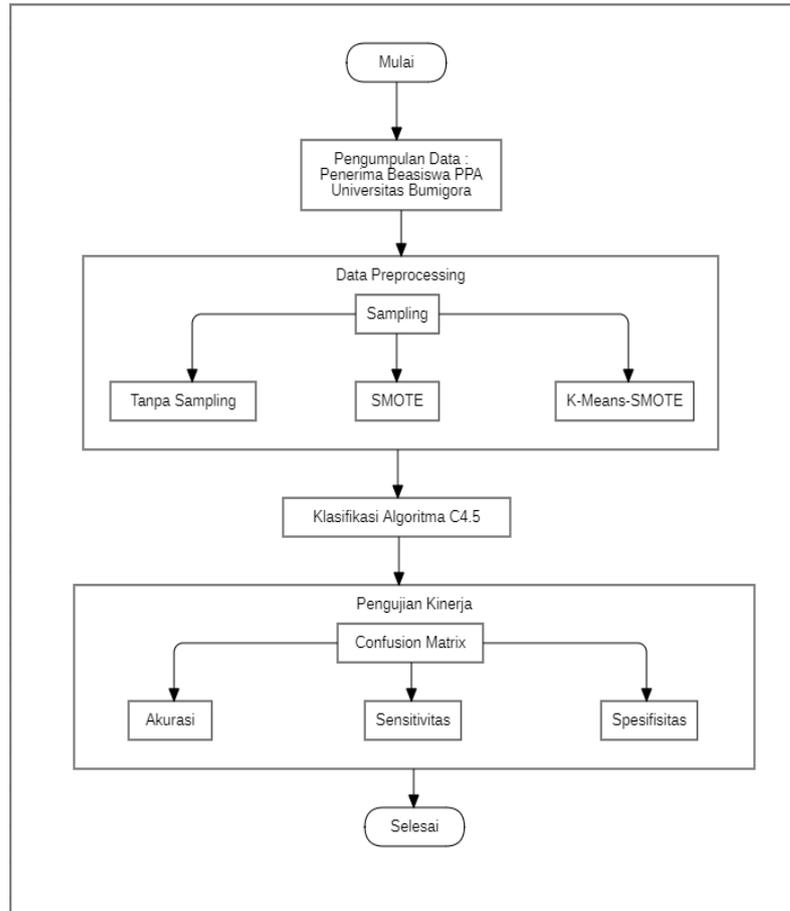
Penelitian [4] menggunakan metode C4.5 untuk klasifikasi penerima beasiswa bidikmisi serta hasil penelitiannya adalah metode C4.5 mendapatkan akurasi sebesar 79.8%. Kelemahan penelitian tersebut, tidak menyelesaikan permasalahan ketidakseimbangan kelas pada data yang digunakan dan tidak hanya itu, penelitian tersebut menggunakan metode pengujian data *split* dengan menentukan jumlah *data training* dan *data testing*, hal tersebut akan berdampak pada hasil klasifikasi yang rawan terjadinya *overfitting*. Penelitian [5] menggunakan metode C4.5 untuk klasifikasi penerima beasiswa. Kelemahan penelitian tersebut, tidak menyelesaikan permasalahan ketidakseimbangan kelas pada data yang digunakan

Berdasarkan kajian penelitian sebelumnya, ada celah penelitian yang bisa dilakukan yaitu belum ada yang mengatasi permasalahan ketidakseimbangan kelas untuk klasifikasi penerima beasiswa. Ketidakseimbangan kelas merupakan jumlah antar kelas pada sebuah data tidak seimbang. Dengan adanya ketidakseimbangan kelas pada data penerima beasiswa dapat mengakibatkan metode klasifikasi cenderung mengklasifikasikan kelas tidak layak (kelas mayoritas) dibandingkan kelas layak (kelas minoritas).

Penelitian ini mengusulkan pendekatan level data seperti metode *SMOTE* [12] dan *k-means-SMOTE* [13] untuk menangani permasalahan ketidakseimbangan kelas pada data beasiswa PPA. Dengan mengatasi permasalahan ketidakseimbangan kelas pada data tersebut, dapat meningkatkan kinerja metode klasifikasi berdasarkan akurasi, sensitivitas, dan spesifisitas. Adapun metode klasifikasi yang digunakan penelitian ini adalah metode C4.5 dikarenakan memiliki kinerja yang baik untuk klasifikasi penerima beasiswa dan termasuk 10 metode paling banyak digunakan [14].

## 2. METODE PENELITIAN

Tahapan penelitian ini menggunakan tahapan *data mining* berdasarkan penelitian [13], [15]. Adapun tahapan penelitiannya terdiri dari pengumpulan data penerima beasiswa PPA Universitas Bumigora, data *preprocessing*, klasifikasi, dan pengujian kinerja menggunakan tabel *confussion matrix*. Adapun detail tahapan penelitian yang digunakan ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian *Data Mining*

Tahapan pertama dalam penelitian ini adalah proses pengumpulan *dataset* penerima beasiswa PPA Universitas Bumigora. *Dataset* tersebut terdiri dari 150 *instance* dan 5 atribut. Dari 150 *instance* terdiri dari 85 *instance* kelas tidak layak dan 65 *instance* kelas layak. Adapun detail atribut *dataset*nya dapat dilihat pada Tabel 1.

Tabel 1. Atribut Data Beasiswa PPA Universitas Bumigora

No	Atribut	Keterangan
1	Jumlah SKS	Jumlah SKS diambil pelamar pada saat proses seleksi
2	Jumlah Tanggungan	Jumlah tanggungan orang tua
3	IPK	Indeks prestasi kumulatif
4	Penghasilan	Penghasilan orang tua
5	Keputusan (Kelas)	Hasil keputusan beasiswa

Tahapan kedua adalah data *preprocessing*. Pada tahapan ini dilakukan *sampling* data pada *dataset* yang digunakan, karena terdapat ketidakseimbangan kelas antara kelas tidak layak (kelas mayoritas) dan 65 kelas layak (kelas minoritas). Penelitian ini mengusulkan menggunakan metode SMOTE dan k-means-SMOTE.

Tahapan ketiga adalah melakukan klasifikasi menggunakan algoritma C4.5. Validasi model metode klasifikasi yang digunakan pada penelitian ini adalah *10-fold cross validation*. Metode *10-fold cross validation* membagi *dataset* menjadi 10 kelompok data.

### 2.1. Metode C4.5

Metode klasifikasi penelitian ini menggunakan metode C4.5. Metode C4.5 salah satu metode klasifikasi dalam *data mining*. Tahapan-tahapan metode C4.5 dalam pembuatan pohon keputusan seperti berikut [16]:

1. Mempersiapkan data latih
2. Perhitungan nilai entropy dengan persamaan (4)

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i \quad (4)$$

3. Perhitungan nilai gain dengan persamaan (5)

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i) \quad (5)$$

4. Mengulangi step 2 sampai seluruh *record* terpartisi.

Tahapan terakhir penelitian ini adalah pengujian kinerja dari algoritma C4.5 dalam mengklasifikasikan penerima beasiswa PPA menggunakan tabel *confusion matrix*. Metode *confusion matrix* menguji 3 hal yaitu akurasi (1), sensitivitas (2), dan spesifisitas (3). Tabel *confusion matrix* dapat dilihat pada Tabel 3.

Tabel 2. *Confussion Matrix*

Kelas Aktual	Kelas Prediksi	
	Tidak Layak (Kelas Mayoritas)	Layak (Kelas Minoritas)
Tidak Layak	<i>True Negative</i> (TN)	<i>False Positive</i> (FP)
Layak	<i>False Negative</i> (FN)	<i>True Positive</i> (TP)

Rumus yang digunakan dalam perhitungan akurasi (1), sensitivitas (2), dan spesifisitas (3) sebagai berikut:

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

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

$$\text{Spesifisitas} = \frac{TN}{TN + FP} \quad (3)$$

*True Positive* (TP) merupakan kelas layak diprediksi secara benar. *False Positive* (FP) merupakan kelas tidak layak diprediksi salah. *True Negative* (TN) merupakan kelas tidak layak diprediksi secara benar. Sedangkan *False Negative* (FN) merupakan kelas layak diprediksi salah.

### 3. HASIL DAN ANALISIS

#### 3.1. Pengumpulan Data

Data penerima beasiswa PPA Universitas Bumigora diperoleh dari bagian kemahasiswaan yaitu wakil rektor 3. Adapun rentang data beasiswa PPA yang digunakan adalah tahun 2015 - 2018. Jumlah data seleksi beasiswa PPA sebanyak 150 *instance*, terdiri dari 85 *instance* kelas tidak layak dan 65 *instance* kelas layak. Adapun contoh datanya ditunjukkan pada Tabel 3.

Tabel 3. Contoh Data Seleksi Beasiswa PPA

No	Jumlah SKS	Jumlah Tanggungan	IPK	Penghasilan Orang Tua	Kelas
1.	22	2	3.72	32.400.000	Tidak Layak
2.	20	3	3.73	46.206.720	Tidak Layak
3.	18	3	3.44	45.013.200	Tidak Layak
4.	24	3	3.5	29.460.000	Tidak Layak
5.	21	1	3.21	14.400.000	Tidak Layak
..	..	..	..	.....	.....
148.	24	3	3.94	1.000.000	Layak
149.	24	4	3.86	1.000.000	Layak
150	24	2	3.47	500.000	Layak

#### 3.2. Data Preprocessing

Pada tahapan ini dilakukan *sampling* data pada *dataset* yang digunakan, karena terdapat ketidakseimbangan kelas antara kelas tidak layak (kelas mayoritas) dan kelas layak (kelas minoritas), hal tersebut akan berdampak pada metode C4.5 yang lebih condong mengklasifikasikan kelas tidak layak dibandingkan kelas layak. Untuk mengatasi permasalahan ketidakseimbangan kelas pada data beasiswa PPA Universitas Bumigora, penelitian ini mengusulkan menggunakan metode SMOTE dan k-means-SMOTE. Adapun rasio kelas *dataset* dapat dilihat pada Tabel 3.

Tabel 4. Rasio Kelas Data Beasiswa PPA Universitas Bumigora

No	Metode <i>samplig</i>	Jumlah <i>Instance</i>	
		Tidak Layak (Kelas Mayoritas)	Layak (Kelas Minoritas)
1	Original	85	65
2	SMOTE	85	85
3	k-means-SMOTE	86	85

### 3.3. Metode Klasifikasi

Metode klasifikasi penelitian ini menggunakan metode C4.5. Adapun tahapan-tahapan perhitungan metode C4.5 seperti berikut:

1. Mempersiapkan data latih. Misalnya dicontohkan *data training* dan *data testing* pada Tabel 4.

Tabel 5. Data Latih

Jumlah SKS	Jumlah Tanggungan	IPK	Penghasilan Orang Tua	Kelas
22	2	3.72	32.400.000	Tidak Layak
20	3	3.73	46.206.720	Tidak Layak
18	3	3.44	45.013.200	Tidak Layak
24	3	3.5	29.460.000	Tidak Layak
21	1	3.21	14.400.000	Tidak Layak
24	2	3.53	750.000	Layak
24	3	3.94	1.000.000	Layak
24	4	3.86	1.000.000	Layak
24	2	3.47	500.000	Layak

Sebelum melakukan perhitungan *entropy* dan *gain*, terlebih dahulu dikonversi *instance* bersifat kontinu pada atribut jumlah sks, jumlah tanggungan, IPK, dan penghasilan orang tua. Adapun hasil konversi atributnya ditunjukkan pada Tabel 5, Tabel 6, Tabel 7, dan Tabel 8.

Tabel 6. Hasil Konversi Atribut Jumlah SKS

Sebelum Konversi	Sesudah Konversi
23 - 24	Banyak
21 - 22	Cukup
≤ 20	Sedikit

Tabel 7. Hasil Konversi Atribut Jumlah Tanggungan

Sebelum Konversi	Sesudah Konversi
≥ 5	Banyak
3 - 4	Sedang
1 - 2	Sedikit

Tabel 8. Hasil Konversi Atribut IPK

Sebelum Konversi	Sesudah Konversi
2.00 - 2.75	Memuaskan
2.76 - 3.50	Sangat Memuaskan
3.51 - 4.00	Dengan Pujian

Tabel 9. Hasil Konversi Atribut Penghasilan Orang Tua

Sebelum Konversi	Sesudah Konversi
≥ 5.000.000	Besar
2.500.000 - 4.999.999	Sedang
1.000.000 - 2.499.000	Cukup
< 1.000.000	Kecil

2. Menghitung nilai *entropy* dengan persamaan (4)

$$Entropy(Total) = \left( -\left(\frac{5}{9}\right) * \left(\log_2\left(\frac{5}{9}\right)\right) \right) + \left( -\left(\frac{4}{9}\right) * \left(\log_2\left(\frac{4}{9}\right)\right) \right) = 0.99107$$

$$Entropy(Jumlah SKS - Banyak) = \left( -\left(\frac{1}{5}\right) * \left(\log_2\left(\frac{1}{5}\right)\right) \right) + \left( -\left(\frac{4}{5}\right) * \left(\log_2\left(\frac{4}{5}\right)\right) \right) = 0.72193$$

$$Entropy(Jumlah SKS - Cukup) = \left( -\left(\frac{2}{2}\right) * \left(\log_2\left(\frac{2}{2}\right)\right) \right) + \left( -\left(\frac{0}{2}\right) * \left(\log_2\left(\frac{0}{2}\right)\right) \right) = 0$$

$$Entropy(Jumlah SKS - Sedikit) = \left( -\left(\frac{2}{2}\right) * \left(\log_2\left(\frac{2}{2}\right)\right) \right) + \left( -\left(\frac{0}{2}\right) * \left(\log_2\left(\frac{0}{2}\right)\right) \right) = 0$$

$$Entropy(Jumlah Tanggungan - Banyak) = \left( -\left(\frac{0}{0}\right) * \left(\log_2\left(\frac{0}{0}\right)\right) \right) + \left( -\left(\frac{0}{0}\right) * \left(\log_2\left(\frac{0}{0}\right)\right) \right) = 0$$

$$Entropy(Jumlah Tanggungan - Sedang) = \left( -\left(\frac{3}{5}\right) * \left(\log_2\left(\frac{3}{5}\right)\right) \right) + \left( -\left(\frac{2}{5}\right) * \left(\log_2\left(\frac{2}{5}\right)\right) \right) = 0.97095$$

$$Entropy(Jumlah Tanggungan - Sedikit) = \left( -\left(\frac{2}{4}\right) * \left(\log_2\left(\frac{2}{4}\right)\right) \right) + \left( -\left(\frac{2}{4}\right) * \left(\log_2\left(\frac{2}{4}\right)\right) \right) = 1$$

3. Menghitung nilai *entropy* dengan persamaan (5)

$$Gain(Jumlah SKS) = 0.99107 - \left( \left(\frac{5}{9}\right) * 0.72192 \right) + \left( \left(\frac{2}{9}\right) * 0 \right) + \left( \left(\frac{2}{9}\right) * 0 \right) = 0.59$$

$$Gain(Jumlah Tanggungan) = 0.99107 - \left( \left(\frac{0}{9}\right) * 0 \right) + \left( \left(\frac{5}{9}\right) * 0.97095 \right) + \left( \left(\frac{4}{9}\right) * 1 \right) = 0.00721$$

Adapun hasil perhitungan *entropy* dan *gain* ditunjukkan pada Tabel 9.

Tabel 10. Hasil Perhitungan *Entropy* dan *Gain* pada Node 1

Atribut	Kelas	Jumlah Kasus (S)	Tidak Layak	Layak	<i>Entropy</i>	<i>Gain</i>
<b>Total</b>		9	5	4	0.99107	
Jumlah SKS	Banyak	5	1	4	0.72192	0.59
	Cukup	2	2	0	0	
	Sedikit	2	2	0	0	
Jumlah Tanggungan	Banyak	0	0	0	0	0.00721
	Sedang	5	3	2	0.97095	
	Sedikit	4	2	2	1	
IPK	Dengan Pujian	5	2	3	0.97095	0.09109
	Sangat Memuaskan	4	3	1	0.81127	
	Memuaskan	0	0	0	0	
Penghasilan Orangtua	Besar	5	5	0	0	<b>0.99107</b>
	Sedang	0	0	0	0	
	Cukup	2	0	2	0	
	Kecil	2	0	2	0	

## 3.4. Pengujian Kinerja

Pengujian kinerja dilakukan untuk mengetahui seberapa baik tingkat akurasi, sensitivitas, dan spesifisitas menggunakan tabel *confusion matrix*. Metode validasi pengujian menggunakan *10-fold cross validation* dengan membagi data sebanyak 10 kelompok. Adapun hasil pengujian ditunjukkan pada Tabel 10 dan Tabel 11.

Tabel 11. *Confusion Matrix* Metode C4.5

Metode <i>sampling</i>	Aktual	Prediksi		Presisi Kelas
		Tidak Layak	Layak	
Tanpa <i>Sampling</i>	Tidak Layak	65	20	76.5%
	Layak	14	51	78.5%
	<i>Recall</i> Kelas	82.3%	71.8%	
SMOTE	Tidak Layak	65	20	76.5%
	Layak	15	70	82.4%
	<i>Recall</i> Kelas	81.3%	77.8%	
k-means-SMOTE	Tidak Layak	66	19	77.6%
	Layak	13	73	84.9%
	<i>Recall</i> Kelas	83.5%	79.3%	

$$\text{Akurasi} = \frac{73 + 66}{73 + 19 + 66 + 13} = \frac{139}{171} = 81.3\%$$

$$\text{Sensitivitas} = \frac{73}{73 + 13} = \frac{73}{86} = 84.9\%$$

$$\text{Spesifisitas} = \frac{66}{66 + 19} = \frac{66}{85} = 77.6\%$$

Tabel 12. Perbandingan Kinerja Metode *Sampling* pada C4.5

Metode <i>sampling</i>	Kinerja		
	Akurasi	Sensitivitas	Spesifisitas
Original	77.3%	78.5%	76.5%
SMOTE	79.4%	82.4%	76.5%
k-means-SMOTE	<b>81.3%</b>	<b>84.9%</b>	<b>77.6%</b>

Berdasarkan Tabel 11, metode k-means-SMOTE memiliki kinerja terbaik berdasarkan akurasi, sensitivitas, dan spesifisitas dibandingkan menggunakan SMOTE dan tanpa *sampling*. Metode k-means-SMOTE memiliki kinerja terbaik, dikarenakan terlebih dahulu melakukan pengelompokan pada kelas mayoritas dan minoritas. Jika pada pengelompokan kelas minoritas terdapat rasio ketidakseimbangan lebih dari 1, maka dilakukan penambahan data pada kelas minoritas menggunakan metode SMOTE [13], [17].

#### 4. KESIMPULAN

Penggunaan pendekatan level data seperti metode SMOTE dan k-means-SMOTE untuk menyelesaikan permasalahan ketidakseimbangan kelas pada data beasiswa PPA dapat meningkatkan kinerja metode C4.5 seperti akurasi, sensitivitas, dan spesifisitas. Adapun kinerja terbaik dihasilkan dari kombinasi metode C4.5 dan k-means-SMOTE dibandingkan metode SMOTE dengan selisih 1.99% untuk akurasi, 2.5% untuk sensitivitas, dan 1.1% untuk spesifisitas. Dengan demikian, penggunaan pendekatan level data (k-means-SMOTE) dan metode klasifikasi C4.5 dapat digunakan untuk menentukan calon penerima beasiswa PPA Universitas Bumigora, karena memiliki kinerja yang baik berdasarkan akurasi, sensitivitas, dan spesifisitas. Untuk penelitian selanjutnya dapat menggunakan pendekatan hibrid yaitu menggabungkan level data (SMOTE dan k-means-SMOTE) dengan *ensemble learning* seperti *boosting*, *bagging* untuk meningkatkan kinerja metode klasifikasi C4.5.

#### UCAPAN TERIMA KASIH

Penulis mencucapkan banyak terimakasih kepada kampus universitas Bumigora yang telah mendukung dalam pemberian data calon penerima Beasiswa Mahasiswa, dan teman-teman dosen dan mahasiswa di lingkungan kampus Universitas Bumigora

#### REFERENSI

- [1] Rismayanti, "Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa di STT Harapan Medan," *Media Infotama*, vol. 12, no. 2, pp. 116–120, 2016.
- [2] J. H. Jaman and N. I. P. Astuti, "Melakukan observasi ke tempat yang akan di teliti ( SDN Karawang Kulon," *Techno Xplore*, vol. 3, no. 1, pp. 25–29, 2018.
- [3] A. S. Suweleh, D. Susilowati, and H. Hairani, "Aplikasi Penentuan Penerima Beasiswa Menggunakan Algoritma C4.5," *Jurnal BITE*, vol. 2, no. 1, pp. 12–21, 2020.
- [4] M. S. Juliardi, R. Saptono, and D. E. Cahyani, "Universitas Sebelas Maret Bidikmisi Applicant's Classification using C4. 5 Algorithm," *ITSMART: Jurnal Teknologi dan Informasi*, vol. 6, no. 1, pp.

- 16–23, 2017.
- [5] N. Hijriana and M. Rasyidan, “Penerapan Metode Decision Tree Algoritma C4.5 untuk Seleksi Calon Penerima Beasiswa Tingkat Universitas,” *Sains Dan Teknologi*, vol. 3, pp. 9–13, 2017.
- [6] D. Noviana, Y. Susanti, and I. Susanto, “Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-Nn) Dan Algoritma C4.5,” *Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT*, pp. 79–87, 2019.
- [7] Okfalisa, R. Fitriani, and Y. Vitriani, “The Comparison of Linear Regression Method and K-Nearest Neighbors in Scholarship Recipient,” in *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2018, pp. 194–199.
- [8] D. Kurniadi, E. Abdurachman, H. L. H. S. Warnars, and W. Suparta, “The prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm,” in *IOP Conference Series: Materials Science and Engineering*, 2018, pp. 1–7.
- [9] N. Z. Dina and R. S. Marjianto, “Prediksi Penentuan Penerima Beasiswa Dengan Metode Knearest Neighbour,” in *Jurnal Nasional Informatika dan Teknologi Jaringan*, 2018, vol. 2, no. 2, pp. 135–139.
- [10] W. Suryaningtyas, N. Iriawan, K. Fithriasari, B. S. S. Ulama, I. Susanto, and A. A. Pravitasari, “On the Bernoulli Mixture Model for Bidikmisi Scholarship Classification with Bayesian MCMC,” in *IOP Conf. Series: Journal of Physics: Conf. Series 1090*, 2018, vol. 1090, no. 1, pp. 1–8.
- [11] H. Sulistiani and Y. T. Utami, “Penerapan Algoritma Klasifikasi sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa,” in *SNTI VI-2018 Universitas Trisakti*, 2018, pp. 300–305.
- [12] Hairani, Setiawan, N. Akhmad, Adji, and T. Bharata, “Metode Klasifikasi Data Mining dan Teknik Sampling SMOTE Menangani Class Imbalance Untuk Segmentasi Customer pada Industri Perbankan,” in *Prosiding SNST Fakultas Teknik*, 2016, vol. 1, no. 1, pp. 168–172.
- [13] H. Hairani, K. E. Saputro, and S. Fadli, “K-means-SMOTE untuk menangani ketidakseimbangan kelas dalam klasifikasi penyakit diabetes dengan C4.5, SVM, dan naive Bayes,” *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 2, pp. 89–93, 2020.
- [14] X. Wu *et al.*, “Top 10 algorithms in data mining,” *Knowledge and information systems*, vol. 14, no. 1, pp. 1–37, 2008.
- [15] Hairani and M. Innuddin, “Kombinasi Metode Correlated Naive Bayes dan Metode Seleksi Fitur Wrapper untuk Klasifikasi Data Kesehatan,” *Jurnal Teknik Elektro*, vol. 11, no. 2, pp. 50–55, 2019.
- [16] L. Nurellisa *et al.*, “Analisis Rekomendasi Calon Debitur Motor pada PT. XYZ Menggunakan Algoritma C4.5,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 4, pp. 673–682, 2020.
- [17] G. Douzas, F. Bacao, and F. Last, “Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE,” *Information Sciences*, vol. 465, pp. 1–20, 2018.