

## **PENINGKATAN KINERJA METODE SVM MENGGUNAKAN METODE KNN IMPUTASI DAN *K-MEANS-SMOTE* UNTUK KLASIFIKASI KELULUSAN MAHASISWA UNIVERSITAS BUMIGORA**

**Hairani\*<sup>1</sup>**

<sup>1</sup>Universitas Bumigora

Email: <sup>1</sup>Hairani@universitasbumigora.ac.id

\*Penulis Korespondensi

(Naskah masuk: 02 April 2020, diterima untuk diterbitkan: 19 Juli 2021)

### **Abstrak**

Salah satu permasalahan utama Universitas Bumigora adalah rasio antara mahasiswa yang masuk dengan mahasiswa lulus tepat waktu tidak seimbang, sehingga akan mengakibatkan penurunan penilaian akreditasi dikemudian hari. Salah satu indikator penilaian dalam proses akreditasi adalah rasio kelulusan mahasiswa. Data kelulusan mahasiswa yang tersimpan pada basisdata kampus, tetapi belum dimanfaatkan dengan maksimal. Dengan memanfaatkan data kelulusan mahasiswa dapat mengetahui pattern atau pola-pola mahasiswa yang lulus tepat waktu atau tidak, sehingga dapat minimalisir terjadinya mahasiswa yang drop out. Tidak hanya itu, pengambil keputusan dapat dimudahkan membuat kebijakan secara dini untuk membantu mahasiswa yang berpotensi drop out dan lulus tidak tepat waktu. Solusi yang ditawarkan pada penelitian ini adalah menggunakan teknik data mining. Salah satu metode data mining yang digunakan penelitian ini adalah metode SVM. Adapun tujuan penelitian ini adalah meningkatkan kinerja metode SVM untuk klasifikasi kelulusan mahasiswa Universitas Bumigora menggunakan metode KNN Imputasi dan *k-means-SMOTE*. Penelitian ini terdiri dari beberapa tahapan yaitu pengumpulan data kelulusan mahasiswa, pra-pengolahan seperti penanganan nilai hilang menggunakan metode KNNI, penanganan ketidakseimbangan kelas menggunakan *k-means-SMOTE*, klasifikasi menggunakan metode SVM. Tahapan terakhir adalah pengujian kinerja SVM berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure*. Berdasarkan hasil pengujian yang telah dilakukan, integrasi metode KNNI, *k-means-SMOTE*, dan SVM mendapatkan akurasi 83.9%, sensitivitas 81.3%, spesifisitas 86.6%, dan *f-measure* 83.5%. Penggunaan metode KNNI dan *k-means-SMOTE* dapat meningkatkan kinerja metode SVM berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure*.

**Kata kunci:** kelulusan mahasiswa, KNN imputasi, *k-means-smote*, SVM

## **IMPROVEMENT PERFORMANCE OF SVM METHOD USING KNN IMPUTATION AND *K-MEANS-SMOTE* METHOD FOR GRADUATION CLASSIFICATION OF BUMIGORA UNIVERSITY STUDENTS**

### **Abstract**

One of the main problems of Bumigora University is the ratio between incoming students and students graduating on time is not balanced, so that it will result in a decrease in accreditation assessment in the future. One of the assessment indicators in the accreditation process is the student graduation ratio. Student graduation data stored in the campus database, but has not been maximally utilized. By utilizing graduation data, students can find out patterns or patterns of students who graduate on time or not, so as to minimize the occurrence of students who drop out. Not only that, decision makers can make it easier to make policies early to help students who have the potential to drop out and not graduate on time. The solution offered in this research is to use data mining techniques. One of the data mining methods used in this study is the SVM method. The purpose of this study is to improve the performance of the SVM method for the classification of Bumigora University graduation students using the KNN Imputation and *k-means-SMOTE* methods. This research consists of several stages, namely the collection of student graduation data, pre-processing such as handling missing values using KNNI method, handling class imbalances using *k-means-SMOTE*, classification the SVM method. The last stage is testing SVM performance based on accuracy, sensitivity, specificity, and *f-measure*. Based on the results of test that have been carried out, the integration of the KNNI, *k-means-SMOTE*, and SVM method get an accuracy of 83.9%, sensitivity 81.3%, specificity 86.6%, and *f-measure* 83.5%. The use of KNNI and *k-means-SMOTE*

method can improve the performance of the SVM method based on accuracy, sensitivity, specificity, and *f-measure*.

**Keywords:** students graduation, *knn imputation*, *k-means-smote*, *SVM*

## 1. PENDAHULUAN

Universitas Bumigora merupakan salah satu universitas swasta yang terdapat di Nusa Tenggara Barat. Salah satu permasalahan utama yang dihadapi adalah terjadi peningkatan mahasiswa baru yang ingin belajar di universitas tersebut setiap tahunnya, akan tetapi tidak diimbangi peningkatan mahasiswa yang lulus tepat waktu. Permasalahan tersebut mengakibatkan rasio antara yang masuk dengan yang keluar tidak seimbang, sehingga akan mengakibatkan penurunan penilaian akreditasi dikemudian hari. Salah satu indikator penilaian dalam proses akreditasi adalah rasio kelulusan mahasiswa. Data kelulusan mahasiswa yang tersimpan pada basisdata kampus, tetapi belum dimanfaatkan dengan maksimal. Dengan memanfaatkan data kelulusan mahasiswa dapat mengetahui pola-pola mahasiswa yang lulus tepat waktu atau tidak, sehingga dapat minimalisir terjadinya mahasiswa yang *drop out*. Tidak hanya itu, pengambil keputusan dapat dimudahkan membuat kebijakan secara dini untuk membantu mahasiswa yang berpotensi *drop out* dan lulus tidak tepat waktu.

Salah satu cara untuk menangani permasalahan diatas, dapat menggunakan teknik data mining untuk prediksi dini kelulusan mahasiswa yang tepat waktu atau tidak, sehingga para pengambil keputusan dapat mengambil kebijakan dengan cepat dan tepat. Data mining merupakan proses pengolahan data menjadi informasi sehingga mendapatkan pola-pola dari sekumpulan data yang besar (Han, Kamber and Pei, 2012).

Penelitian klasifikasi kelulusan mahasiswa menggunakan teknik data mining sudah banyak dilakukan. Penelitian (Maesya and Hendiyanti, 2019) menggunakan metode *CART* untuk prediksi kelulusan mahasiswa dengan akurasi sebesar 77.5%. Penelitian (Arifin and Hadiana, 2019) menggunakan metode *C4.5* untuk prediksi kegagalan studi mahasiswa dengan akurasi sebesar 82.52%. Penelitian (Aesyi and Wardoyo, 2019) menggunakan metode *Case Based Reasoning* (CBR) untuk prediksi kelulusan mahasiswa dengan akurasi 76%. Penelitian (Bisri and Rachmatika, 2019) menggunakan gabungan metode *Gradient Boosted Tree* (GBT), *SMOTE*, dan *Bagging* untuk klasifikasi kelulusan mahasiswa dengan akurasi sebesar 80.57%.

Penelitian (Fadli, Zulfa and Ramadhani, 2018) membandingkan metode *C4.5*, *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) untuk klasifikasi kelulusan mahasiswa. Berdasarkan hasil penelitiannya, metode SVM

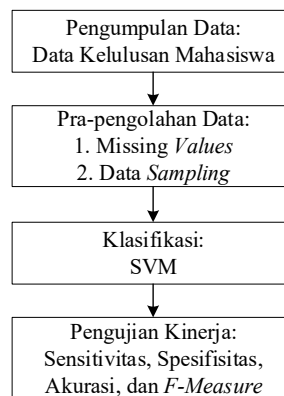
memiliki akurasi terbaik dibandingkan metode ANN dan *C4.5* sebesar 90.55%. Penelitian (Rahman and Davis, 2013) melakukan pra-pengolahan data penyakit jantung menggunakan metode *mean* untuk menangani nilai hilang dan metode *Under-Sampling* menangani data tidak seimbang. Sedangkan metode klasifikasi yang digunakan adalah metode *FURIA* dan *Decision Tree*.

Berdasarkan kajian literatur diatas terdapat persamaan serta perbedaan penelitian ini dengan penelitian (Rahman and Davis, 2013). Persamaannya terletak pada proses pra-pengolahan seperti penanganan nilai hilang dan data tidak seimbang. Adapun perbedaannya adalah penggunaan metode yang berbeda dalam menangani permasalahan nilai hilang dan data tidak seimbang pada data yang digunakan. Penelitian ini menggunakan metode *KNN Imputasi* untuk menangani permasalahan nilai hilang pada data kelulusan mahasiswa. Metode *k-means-SMOTE* digunakan untuk menangani ketidakseimbangan kelas pada data kelulusan mahasiswa Universitas Bumigora, karena memiliki kelas lulus tidak tepat waktu lebih banyak dibandingkan lulus tepat waktu. Hal tersebut akan berdampak pada kecenderungan metode klasifikasi untuk mengklasifikasikan kelas lulus tidak tepat waktu dibandingkan lulus tepat waktu. Sedangkan metode klasifikasi yang digunakan adalah metode SVM.

Penggabungan metode *KNNI* dan *k-means-SMOTE* pada proses pra-pengolahan bertujuan untuk meningkatkan kinerja metode SVM berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure* untuk klasifikasi kelulusan mahasiswa Universitas Bumigora.

## 2. METODE PENELITIAN

Alur penelitian ini mengacu pada Gambar 1.



Gambar 1. Alur Penelitian

## 2.1. Pengumpulan Data

Pengumpulan data dilakukan di Universitas Bumigora. Data yang diambil adalah data kelulusan mahasiswa angkatan 2009-2012 berjumlah 302 data dan memiliki 8 atribut. Atribut data kelulusan yang digunakan penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Atribut Data Kelulusan Mahasiswa Universitas Bumigora

No	Atribut	Keterangan	Tipe Data
1.	Jenis Kelamin	Jenis Kelamin	Nominal ( Laki, Perempuan )
2.	IPS 1	Indeks Prestasi Semester 1	Numerik
3.	IPS 2	Indeks Prestasi Semester 2	Numerik
4.	IPS 3	Indeks Prestasi Semester 3	Numerik
5.	IPS 4	Indeks Prestasi Semester 4	Numerik
6.	IPS 5	Indeks Prestasi Semester 5	Numerik
7.	IPK	Indeks Prestasi Kumulatif	Numerik
8.	Status Kelulusan	Kelas (Variabel Target )	Nominal (Tepat Waktu, Tidak Tepat Waktu)

## 2.2. Pra-Pengolahan Data

Pra-pengolahan data diperlukan untuk menjamin kualitas data yang digunakan pada proses klasifikasi, agar metode SVM mendapatkan kinerja yang optimal. Pra-pengolahan yang digunakan pada penelitian ini adalah penanganan nilai hilang menggunakan metode KNNI dan data tidak seimbang menggunakan metode *k-means-SMOTE*. Data kelulusan mahasiswa yang digunakan memiliki 7 atribut masukan dan 1 atribut target, dan jumlah datanya sebanyak 302 *instance*.

### 2.2.1. Nilai Hilang

Nilai hilang merupakan sebuah nilai atau informasi yang hilang dari data dan dapat menurunkan akurasi metode klasifikasi (Hartini, 2017). Data kelulusan yang digunakan sebanyak 302 *instance*, terdapat nilai hilang sebanyak 94 *instance*. Penelitian ini menggunakan metode KNNI untuk menangani data nilai hilang, karena memiliki kinerja yang baik (Hartini, 2017). Metode KNNI merupakan sebuah metode yang mengimplementasikan konsep kerja metode *Nearest Neighbor*. Metode KNNI menangani permasalahan nilai hilang dengan cara mengisi data yang terdapat nilai hilang dengan nilai tetangga terdekat berdasarkan *k* diberikan. Data yang terdapat nilai hilang ditunjukkan pada Tabel 2, dan data nilai hilang yang sudah ditangani menggunakan metode KNNI ditunjukkan pada Tabel 3.

Tahapan-tahapan metode KNNI (Hartini, 2017) (Minakshi, Vohra and Gimpny, 2014):

1. Tentukan nilai *K*, Penelitian ini menggunakan *K* = 5.
2. Menghitung jarak antar *instance* nilai hilang dan *instance* data training. Formula digunakan untuk menghitung jarak ditunjukkan pada persamaan (1).

$$D(x, y) = \sum_{i=1}^n \sqrt{x_i^2 - y_i^2} \quad (1)$$

3. Setelah menghitung jarak dipilih nilai data yang memiliki jarak terdekat. Jika menggunakan nilai *K*=5, maka akan dipilih 5 nilai dengan jarak terdekat. Hitung rata-rata dari nilai *K*=5 yang terdekat.
4. Menghitung nilai rata-rata berdasarkan *K* yang dipilih. Formulasi yang digunakan untuk menghitung rata-ratanya ditunjukkan pada persamaan (2).

$$M = \frac{1}{n} \sum_{i=1}^n m_i \quad (2)$$

5. Mengisi nilai hilang dengan hasil dari proses 4.

Tabel 2. Data Nilai Hilang

No.	JK	IP Semester					IPK	Status
		1	2	3	4	5		
1.	L	2.55	2.52	3	2.93	2.99	3.02	TTW
2.	P	3.16	3.1	3	3.29	3.33	3.35	TW
3.	L	2.01	2.12	2	2.4	2.51	2.64	TTW
4.	L	1.62	1.7	2	2.01	2.14	2.41	TTW
5.	L	2.94	2.86	3	3.17	3.19	3.2	TTW
6.	L	2.31	2.31	2	Nan	Nan	Nan	TTW
7.	L	2.63	2.73	3	2.91	2.97	3.03	TW
8.	P	2.26	2.26	2	Nan	Nan	Nan	TTW
9.	L	2.34	2.36	3	2.59	2.69	2.76	TTW
10.	L	3.35	3.47	3	3.46	3.46	Nan	TW
11.	L	2.07	2.22	2	2.4	2.32	2.32	TTW
12.	P	3.42	2.85	3	3.13	3.5	3.15	TW
13.	L	3.17	2.9	3	3.39	3.41	3.22	TW
14.	L	Nan	2.32	2	2.5	2.5	2.42	TTW
15.	P	3.42	2.85	3	3.13	3.5	3.15	TW

Tabel 3. Hasil Penanganan Nilai Hilang Menggunakan KNN Imputasi

No.	JK	IP Semester					IPK	Status
		1	2	3	4	5		
1.	L	2.55	2.52	3	2.93	2.99	3.02	TTW
2.	P	3.16	3.1	3	3.29	3.33	3.35	TW
3.	L	2.01	2.12	2	2.4	2.51	2.64	TTW
4.	L	1.62	1.7	2	2.01	2.14	2.41	TTW
5.	L	2.94	2.86	3	3.17	3.19	3.2	TTW
6.	L	2.31	2.31	2	2.38	2.43	2.5	TTW
7.	L	2.63	2.73	3	2.91	2.97	3.03	TW
8.	P	2.26	2.26	2	2.38	2.43	2.5	TTW
9.	L	2.34	2.36	3	2.59	2.69	2.76	TTW
10.	L	3.35	3.47	3	3.46	3.46	3.21	TW
11.	L	2.07	2.22	2	2.4	2.32	2.32	TTW
12.	P	3.42	2.85	3	3.13	3.5	3.15	TW
13.	L	3.17	2.9	3	3.39	3.41	3.22	TW
14.	L	2.07	2.32	2	2.5	2.5	2.42	TTW
15.	P	3.42	2.85	3	3.13	3.5	3.15	TW

### 2.2.2. Data Sampling

Data kelulusan yang digunakan penelitian ini terdapat ketidakseimbangan kelas antara kelas tepat waktu dan tidak tepat waktu. Penelitian ini menggunakan metode *k-means-SMOTE* untuk menangani ketidakseimbangan kelasnya. *k-means-SMOTE* merupakan hasil modifikasi dari metode *synthetic minority oversampling technique* (SMOTE) yang dikembangkan oleh (Douzas, Bacao and Last, 2018). Metode Smote menambahkan kelas minoritas

dengan cara membangkitkan data sintesis berdasarkan *k-nearest neighbor*) antar kelas minoritas (Chawla, Bowyer and Hall, 2002). Sedangkan pada metode *k-means-SMOTE* terlebih dahulu melakukan pengelompokan pada kelas mayoritas dan minoritas. Jika suatu kelompok terdapat nilai rasio ketidakseimbangan  $> 1$ , maka kelas minoritas ditambahkan dengan metode SMOTE (Douzas, Bacao and Last, 2018). Data kelulusan yang tidak seimbang dan juga data seimbang menggunakan metode *k-means-SMOTE* ditunjukkan pada Tabel 4. Adapun Tahapan-tahapan penanganan ketidakseimbangan kelas menggunakan *k-means-SMOTE* terdiri dari 3 tahapan yaitu pengelompokan menggunakan K-means, filter, dan *oversampling* menggunakan algoritma Smote (Douzas, Bacao and Last, 2018; Hairani, Saputro and Fadli, 2020).

Tabel 4. Perbedaan Kelas Data Kelulusan Mahasiswa

Metode Sampling	Jumlah Instance		
	Tepat Waktu	Tidak Waktu	Tepat
Orginial	78	224	
<i>k-means-SMOTE</i>	224	224	

### 2.3. Metode Klasifikasi

Metode klasifikasi yang digunakan penelitian ini adalah metode *Support Vector Machine* (SVM). Metode SVM merupakan sebuah metode klasifikasi *supervised learning* yang bertujuan untuk mencari atau menemukan *hyperplane* yang optimal dengan memaksimalkan jarak atau *margin* antar kelas data dengan menggunakan persamaan (3) (Zaki, Meira Jr and Meira, 2014).

$$h(x) = w^T x + b \quad (3)$$

$$= w_1 x_1 + w_2 x_2 + \dots + w_d x_d + b$$

$W$  merupakan bobot, sedangkan  $b$  adalah bias. Persamaan (3) dapat didetailkan menjadi beberapa persamaan yang ditunjukkan pada persamaan (4) dan (5).

$$\left[ (w^T \cdot x) + b \right] \geq 1 \text{ untuk } y_i = +1 \quad (4)$$

$$\left[ (w^T \cdot x) + b \right] \geq -1 \text{ untuk } y_i = -1 \quad (5)$$

Metode SVM tidak hanya bekerja pada data bersifat linier tetapi juga bekerja pada data yang non linier. Pada kenyataannya, tidak semua data bersifat non-linear dapat ditransformasikan ke bentuk linier. Permasalahan tersebut dapat diatasi dengan menggunakan dua pendekatan yaitu *soft margin hyperplane* dan *feature space*. Pendekatan *soft margin hyperplane* digunakan untuk merubah data bersifat non-linear menjadi linier dengan garis atau bidang tetap yang bersifat fleksibel. Pada *soft margin hyperplane* menggunakan formulasi variabel *slack*  $\xi$  yang ditunjukkan pada persamaan (6).

$$\begin{aligned} x_i \cdot w + b &\geq 1 - \xi \text{ untuk } y_i = \text{kelas 1} \\ x_i \cdot w + b &\leq -1 + \xi \text{ untuk } y_i = \text{kelas 2} \end{aligned} \quad (6)$$

Adapun formulasi untuk menemukan bidang pemisah menggunakan persamaan (7).

$$\min \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^n \xi_i \right) \quad (7)$$

Nilai  $C$  merupakan parameter pinalti yang ditentukan pengguna untuk toleransi kesalahan dalam klasifikasi data.

Sedangkan pendekatan *feature space* digunakan untuk transformasi dari ruang *input* yang hanya memisahkan data linier ke bentuk dimensi tinggi. Proses perubahan ruang *input* menjadi dimensi tinggi membutuhkan waktu komputasi yang lama. Oleh karena itu, fungsi kernel dibutuhkan agar mempercepat waktu komputasi.

Adapun macam-macam fungsi kernel pada metode SVM yang dapat digunakan ditunjukkan pada persamaan (8), (9), (10), dan (11) (Prasetyo, 2014).

#### 1. Kernel Linier

$$K(x_i, x) = x_i^T x \quad (8)$$

#### 2. Kernel Gaussian Radial Basis Function (RBF)

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (9)$$

#### 3. Kernel Polynomial

$$K(x_i, x_j) = ((x_i \cdot x_j) + c)^d \quad (10)$$

#### 4. Sigmoid kernel

$$K(x_i, x_j) = \tanh(\gamma x_i^T x + r) \quad (11)$$

### 2.4. Pengujian Kinerja

Pengujian kinerja menggunakan tabel konfusi. Tabel konfusi digunakan menghitung jumlah data yang diklasifikasikan secara benar dan salah yang ditunjukkan pada Tabel 5.

Tabel 5. Matrik Konfusi

Aktual	Prediksi	
	Tepat Waktu	Tidak Tepat Waktu
Tepat Waktu	TP	FN
Tidak Tepat Waktu	FP	TN

Formula digunakan untuk menghitung akurasi (12), Sensitivitas (13), Spesifisitas (14), Presisi (15), dan *F-measure* (16) (Luque, et al., 2019).

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + TN + FP} \quad (12)$$

$$\text{Sensitivitas (Recall)} = \frac{TP}{TP + FN} \quad (13)$$

$$\text{Spesifisitas} = \frac{TN}{TN + FP} \quad (14)$$

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (15)$$

$$F - measure = 2 \left( \frac{\text{Presisi} * \text{Recall}}{\text{Presisi} + \text{Recall}} \right) \quad (16)$$

*True Positive* (TP) merupakan kelas tepat waktu yang diprediksi secara benar. *False Positive* (FP) merupakan kelas tidak tepat waktu, tetapi diprediksi tepat waktu. *True Negative* (TN) merupakan kelas tidak tepat waktu yang diprediksi secara benar. *False Negative* (FN) merupakan kelas tepat waktu, tetapi diprediksi tidak tepat waktu.

### 3. HASIL DAN PEMBAHASAN

Penelitian ini dimulai dari tahapan pengumpulan data, pra-pengolahan data, klasifikasi, dan pengujian kinerja. Data yang digunakan penelitian ini adalah data kelulusan mahasiswa Universitas Bumigora angkatan 2009-2012. Pra-pengolahan penelitian ini menggunakan metode KNNI untuk penanganan nilai hilang, *k-means-SMOTE* digunakan untuk menangani permasalahan ketidakseimbangan kelas pada data kelulusan yang digunakan, dan metode SVM digunakan sebagai metode klasifikasi. Pengujian kinerja menggunakan tabel konfusi berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure*. Berdasarkan pengujian metode SVM yang telah dilakukan menggunakan *10-fold cross validation*, didapatkan hasil yang ditunjukkan pada Tabel 6 dan Tabel 7.

Tabel 6. Matrik Konfusi Metode SVM

Prediksi	Aktual			
	Original		<i>k-means-SMOTE</i>	
	Tepat Waktu	Tidak Tepat Waktu	Tepat Waktu	Tidak Tepat Waktu
Tepat Waktu	33	45	182	42
Tidak Tepat Waktu	20	204	30	194

$$\text{Akurasi} = \frac{182 + 194}{182 + 30 + 194 + 42} = 83.9\%$$

$$\text{Sensitivitas} = \frac{182}{182 + 42} = 81.3\%$$

$$\text{Spesifisitas} = \frac{194}{194 + 30} = 86.6\%$$

$$\text{Presisi} = \frac{182}{182 + 30} = 85.8\%$$

$$F - measure = 2 \left( \frac{81.3 * 85.8}{81.3 + 85.8} \right) = 83.5\%$$

Tabel 7. Hasil Akurasi Metode SVM

Kinerja	SVM	SVM + KNNI + K-Means-Smote (Metode usulan)
Akurasi	78.5%	83.9%
Sensitivitas	42.3%	81.3%
Spesifisitas	91.1%	86.6%
<i>F-measure</i>	50.4%	83.5%

Berdasarkan Tabel 7 menunjukkan peningkatan kinerja metode SVM dengan KNNI dan *k-means-*

*SMOTE* berdasarkan akurasi, sensitivitas, dan *f-measure*. Kenaikan kinerja sensitivitas memiliki peranan sangat penting untuk meningkatkan kinerja akurasi dan *f-measure*. Tanpa *sampling*, metode SVM hanya mengklasifikasikan kelas tidak tepat waktu dan mengabaikan kelas tepat waktu sehingga kinerja sensitivitasnya menjadi paling rendah. Kinerja sensitivitas tanpa *sampling* sebesar 42.3%, sedangkan menggunakan *oversampling* *k-means-SMOTE* sebesar 81.3%. Terjadi kenaikan sebesar 39% pada kinerja sensitivitas, dikarenakan ada penambahan data kelas tepat waktu menggunakan *k-means-SMOTE*. Hal tersebut dapat meningkatkan metode SVM untuk klasifikasi kelas tepat waktu secara benar, sehingga kinerja sensitivitasnya naik. Dengan demikian, penggunaan metode KNNI dan *k-means-SMOTE* dapat meningkatkan kinerja metode SVM (Hartini, 2017)(Capariño, Sison and Medina, 2019)(Jordanov, Petrov and Petrozziello, 2018) (Douzas, Bacao and Last, 2018; Hairani, Saputro and Fadli, 2020).

### 4. KESIMPULAN

Metode yang diusulkan (KNNI - *k-means-SMOTE* - SVM) menghasilkan kinerja terbaik dibandingkan tanpa *sampling* berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure* masing-masing sebesar 83.9%, 81.3%, 86.6%, dan 83.5%. Terjadi peningkatan sebesar 5.4% untuk akurasi, 39% untuk sensitivitas, dan 33.1% untuk *f-measure* karena penambahan data kelas tepat waktu menggunakan *k-means-SMOTE*. Dengan demikian, metode yang diusulkan dapat digunakan untuk klasifikasi kelulusan mahasiswa Universitas Bumigora.

### DAFTAR PUSTAKA

- AESYI, U. S. AND WARDOYO, R. (2019) 'Prediction of Length of Study of Student Applicants Using Case Based Reasoning', *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(1), p. 11. doi: 10.22146/ijccs.28076.
- ARIFIN, D. AND HADIANA, A. (2019) 'Computer-based Techniques for Predicting the Failure of Student Studies Using the Decision Tree method', *IOP Conference Series: Materials Science and Engineering*, 662(2), pp. 1–9. doi: 10.1088/1757-899X/662/2/022112.
- BISRI, A. AND RACHMATIKA, R. (2019) 'Integrasi Gradient Boosted Trees dengan SMOTE dan Bagging untuk Deteksi Kelulusan Mahasiswa', *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, 8(4), p. 309. doi: 10.22146/jnteti.v8i4.529.
- CAPARIÑO, E. T., SISON, A. M. AND MEDINA,

- R. P. (2019) 'Application of the modified imputation method to missing data to increase classification performance', in *2019 IEEE 4th International Conference on Computer and Communication Systems, ICCCS 2019*. IEEE, pp. 134–139. doi: 10.1109/CCOMS.2019.8821632.
- CHAWLA, N. V, BOWYER, K. W. AND HALL, L. O. (2002) 'SMOTE: Synthetic Minority Over-sampling TEchnique', *Journal of Artificial Intelligence Research*, 16, pp. 341–378.
- DOUZAS, G., BACAO, F. AND LAST, F. (2018a) 'Improving Imbalanced Learning Through A Heuristic Oversampling Method Based On K-means and SMOTE', *Information Sciences*. Elsevier Inc., 465, pp. 1–20. doi: 10.1016/j.ins.2018.06.056.
- FADLI, A., ZULFA, M. I. AND RAMADHANI, Y. (2018) 'Performance Comparison of Data Mining Classification Algorithms for Early Warning System of Students Graduation Timeliness', *Jurnal Teknologi dan Sistem Komputer*, 6(4), p. 158. doi: 10.14710/jtsiskom.6.4.2018.158-163.
- HAIRANI, H., SAPUTRO, K. E. AND FADLI, S. (2020) 'K-means-SMOTE untuk menangani ketidakseimbangan kelas dalam klasifikasi penyakit diabetes dengan C4.5, SVM, dan naive Bayes', *Jurnal Teknologi dan Sistem Komputer*, 8(2), pp. 89–93. doi: <https://doi.org/10.14710/jtsiskom.8.2.2020.89-93>.
- HAN, J., KAMBER, M. AND PEI, J. (2012) 'Data Mining Concepts and Techniques', in. Waltham: Morgan Kaufmann.
- HARTINI, E. (2017) 'Implementation of Missing Values Handling Method for Evaluating the System/Component Maintenance Historical Data', *Jurnal Teknologi Reaktor Nuklir Tri Dasa Mega*, 19(1), p. 11. doi: 10.17146/tdm.2017.19.1.3159.
- JORDANOV, I., PETROV, N. AND PETROZZIELLO, A. (2018) 'Classifiers Accuracy Improvement Based on Missing Data Imputation', *Journal of Artificial Intelligence and Soft Computing Research*, 8(1), pp. 31–48. doi: 10.1515/jaiscr-2018-0002.
- LUQUE, A. ET AL. (2019) 'The impact of class imbalance in classification performance metrics based on the binary confusion matrix', *Pattern Recognition*. Elsevier Ltd, 91, pp. 216–231. doi: 10.1016/j.patcog.2019.02.023.
- MAESYA, A. AND HENDIYANTI, T. (2019) 'Forecasting Student Graduation with Classification and Regression Tree (CART) Algorithm', *IOP Conference Series: Materials Science and Engineering*, 621(1), pp. 1–6. doi: 10.1088/1757-899X/621/1/012005.
- MINAKSHI, VOHRA, R. AND GIMPY (2014) 'Missing Value Imputation in Multi Attribute Data Set', *International Journal of Computer Science and Information Technologies*, 5(4), pp. 5315–5321.
- PRASETYO, E. (2014) *Data Mining, Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi.
- RAHMAN, M. M. AND DAVIS, D. N. (2013) 'Cluster Based Under-Sampling for Unbalanced Cardiovascular Data', in *Proceedings of the World Congress on Engineering, Vol. 3*.
- ZAKI, M. J., MEIRA JR, W. AND MEIRA, W. (2014) *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.