

PERBANDINGAN ALGORITME *MACHINE LEARNING* UNTUK MEMPREDIKSI PENGAMBIL MATA KULIAH

Fitra A. Bachtiar¹, Indra K. Syahputra², Satrio A. Wicaksono³

¹Fakultas Ilmu Komputer, Universitas Brawijaya Malang
Email: ¹fitra.bachtiar@ub.ac.id, ²kurniawan.isyahputra@gmail.com, ³satrio@ub.ac.id

(Naskah masuk: 05 Februari 2019, diterima untuk diterbitkan: 18 Februari 2019)

Abstrak

Pada setiap awal semester bagian akademik melakukan penjadwalan dan penentuan matakuliah yang akan dibuka untuk semester berikutnya. Akan tetapi proses tersebut memiliki permasalahan antara lain kelas yang dibuka terlalu banyak dibanding jumlah siswa yang berminat atau sebaliknya. Selain itu, dalam permasalahan prediksi data yang terkumpul memiliki kecenderungan tidak seimbang pada setiap kelas (*imbalance class*). Hal ini akan berdampak pada proses penjadwalan yang kurang tepat. Sehingga dibutuhkan sistem yang dapat memprediksi mahasiswa pengambil mata kuliah. Akan tetapi ada banyak algoritme yang dapat digunakan untuk proses prediksi. Penelitian ini membandingkan performa algoritma untuk klasifikasi mahasiswa pengambil matakuliah. Pada penelitian ini prediksi dilakukan berdasarkan atribut dari data mahasiswa. Atribut-atribut tersebut yaitu Nilai, IP, IPK, SKS, SKSK dan Semester. Pada setiap observasi pada atribut-atribut tersebut prediksi akan dilakukan apakah mahasiswa tersebut mengambil mata kuliah tertentu. Prediksi dibagi menjadi 2 kelas yaitu 'Ya' untuk mahasiswa yang diprediksi mengambil matakuliah dan 'Tidak' untuk mahasiswa yang diprediksi tidak mengambil matakuliah. Teknik *Synthetic Minority Oversampling Technique* (SMOTE) digunakan untuk menangani data yang tidak seimbang. Pada penelitian ini klasifikasi dilakukan dengan membandingkan algoritme *k-Nearest Neighbor* (*k-NN*) dan *Support Vector Machine* (SVM) untuk kasus prediksi pengambil matakuliah. Hasil pengujian menggunakan 3 mata kuliah sebagai sampel. Dari hasil rerata, diperoleh hasil prediksi *k-NN* memiliki kinerja yang lebih baik daripada SVM. Selain itu, penggunaan teknik SMOTE dapat mempengaruhi hasil klasifikasi berupa peningkatan nilai AUC, CA, F1, *precision* dan *recall*.

Kata kunci: *knn, svm, smote, klasifikasi, data mining, mahasiswa*

Abstract

At the beginning of each semester, the academic section conducts scheduling and determining the courses offered for the next semester. However, the process has problems such as too many classes offered to the student compared to the number of students who take the class or vice versa. Besides that, in the prediction problems, the collected data has an imbalance tendency in each class. As a result, these problems could cause in ineffective scheduling. Thus, there is a need to build a system that can predict students taking courses. However, there are many algorithms that can be used for the prediction. This study compares the performance of algorithms for classifications of students taking courses. In this study, predictions are modeled based on the attributes of student data, namely Grades, GPA, Cumulative GPA, Semester Credits, Cumulative Semester Credits and Semester. The classification process will be carried out to produce a prediction of whether the student takes a particular subject or not. Classification results are divided into 2 classes, namely 'Yes' for students who are predicted to take and 'No' for students who are predicted not to take the class. To handle imbalance dataset will use Synthetic Minority Oversampling Technique (SMOTE) techniques. Classification method used in this study are k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) algorithms to compare their performance for prediction cases. The test results used 3 courses as a sample. In average k-NN prediction results have a better performance than SVM. In addition, the use of SMOTE techniques can influence the classification results in the form of an increase in AUC, CA, F1, precision and recall values.

Keywords: *comparison, knn, svm, data mining, student*

1. PENDAHULUAN

Perguruan tinggi merupakan institusi pendidikan sebagai penyelenggara kegiatan

akademik bagi mahasiswa. Penyelenggaraan kegiatan akademik tersebut diharapkan mampu memberikan pelayanan dan menyelenggarakan pendidikan yang berkualitas bagi mahasiswa. Salah satu bagian dalam perguruan tinggi yang memiliki pengaruh terhadap

proses didalam penyelenggaraan pendidikan tersebut adalah bagian akademik. Bagian akademik memiliki beberapa tugas, salah satu tugas yang penting dan memerlukan perhatian khusus bagian akademik dalam perguruan tinggi adalah saat masa pengisian Kartu Rencana Studi (KRS) mahasiswa. Pada masa ini, sebelum kegiatan belajar mengajar semester dimulai bagian akademik tiap fakultas memiliki kewajiban untuk merancang jadwal mata kuliah sebelum dapat diambil oleh mahasiswa.

Akan tetapi proses merancang jadwal mata kuliah cukup rumit dan seringkali menimbulkan beberapa permasalahan. Permasalahan tersebut diantaranya adalah kelas yang tersedia tidak dapat memenuhi jumlah mahasiswa yang berminat untuk mengambil kelas tersebut dan kelas yang tersedia cukup banyak namun sepi peminat. Akibatnya, pihak akademik selalu menambah waktu pengisian KRS untuk mengakomodasi mahasiswa yang tidak dapat memprogram kelas pada KRS mahasiswa tersebut. Hal ini dapat berakibat mundurnya jadwal kegiatan perkuliahan. Selain itu, beberapa matakuliah yang diprogram oleh mahasiswa adalah matakuliah pilihan. Berbeda dengan matakuliah wajib yang wajib diambil. Jumlah mahasiswa yang mengambil matakuliah wajib relatif lebih banyak dibandingkan dengan matakuliah pilihan. Hal tersebut mengakibatkan data matakuliah yang ada memiliki ketidak seimbangan data. Hal-hal tersebut berakibat pada proses penjadwalan yang kurang efisien. Dari uraian permasalahan tersebut maka dibutuhkan suatu sistem yang dapat membantu bagian akademik untuk melakukan prediksi apakah mahasiswa mengambil mata kuliah atau tidak.

Terdapat beberapa penelitian terkait dengan prediksi dan pendidikan. Penelitian sebelumnya telah menggunakan *NBTree* untuk memprediksi performa mahasiswa (Christian, 2014). Atribut yang digunakan pada penelitian tersebut adalah jenis kelamin, fakultas, jurusan, jalur masuk, nilai ujian, kota, jurusan SMA, SKS total, IPK, dan waktu studi. Penelitian lain melakukan data mining dengan menggunakan *Naive Bayes* untuk memprediksi performa mahasiswa (Pujiyanto, 2017). Fitur data yang digunakan dalam penelitian tersebut adalah nilai Ujian Nasional, nilai Ujian Nasional Bahasa Inggris, nilai Ujian Nasional Bahasa Indonesia, jumlah buku yang dibaca perhari, prestasi dan IPK. Selain itu, metode *k-Nearest Neighbor (k-NN)* digunakan untuk memprediksi kemampuan siswa dalam belajar (Taruna, 2014). Dalam penelitian lain, *Support Vector Machine (SVM)* telah diaplikasikan untuk untuk mendiagnosa siswa dalam kesulitan belajar (Wu, 2008). Metode klasifikasi *k-NN* memiliki karakteristik klasifikasi dengan mempertimbangkan tetangga terdekat, sedangkan metode klasifikasi SVM bekerja dengan baik pada data yang secara linier dapat dipisahkan (Octaviani, 2014).

Tujuan dari penelitian ini adalah mengimplementasikan dan membandingkan metode

klasifikasi *data mining* yaitu SVM, dan *k-NN* untuk melakukan klasifikasi terhadap mahasiswa pengambil mata kuliah. Atribut yang digunakan untuk klasifikasi didapat dari atribut akademik mahasiswa. Atribut akademik tersebut yaitu Nilai, IP, IP Kumulatif, SKS, SKS Kumulatif dan Semester. Tahap *pre-processing* menggunakan metode *Synthetic Minority Over-Sampling Technique (SMOTE)* dalam menangani masalah *dataset* yang tidak seimbang (*imbalance*) dan menggunakan teknik *Cross Validation* pada tahap pengujian.

2. DASAR TEORI

2.1 Support Vector Machine (SVM)

SVM adalah suatu metode yang dapat digunakan untuk prediksi, baik dalam kasus klasifikasi maupun regresi (Santosa, 2007). Pada dasarnya SVM merupakan linier classifier yang dapat mengklasifikasikan secara linier. Akan tetapi, SVM telah dikembangkan untuk dapat menyelesaikan permasalahan non-linier dengan menggunakan konsep kernel dengan mentransformasikan ke ruang berdimensi tinggi. SVM memetakan masalah yang tidak dapat diselesaikan dalam ruang dimensi linier ke ruang nonlinier berdimensi tinggi (Hou, 2017). Proses penyelesaian tersebut adalah dengan mencari *hyperplane* yang dapat memaksimalkan jarak (*margin*) antara kelas data, yang dinotasikan sebagai:

$$f(x) = w^T x + b \quad (1)$$

Dari persamaan (1), menurut Vapnik dan Cortes (1995) dapat diperoleh persamaan :

$$[(w^T \cdot x_i) + b] \geq 1 \text{ untuk } y_i = +1 \quad (2)$$

$$[(w^T \cdot x_i) + b] \leq -1 \text{ untuk } y_i = -1 \quad (3)$$

SVM dapat bekerja pada data non-linier dengan menggunakan pendekatan kernel yang diterapkan pada fitur data awal himpunan data. Fungsi kernel pada SVM adalah untuk memetakan dimensi awal, pada umumnya dimensi fitur yang lebih rendah, pada himpunan data ke dimensi baru yang memiliki dimensi yang relatif lebih tinggi. Beberapa macam fungsi kernel (Prasetyo, 2012) diantaranya :

1. Kernel Gaussian Radial Basis Function (RBF)

$$K(x_i, x_j) = \exp\left(\frac{\|x_i - x_j\|^2}{2a^2}\right) \quad (4)$$

2. Kernel Polynomial

$$K(x_i, x_j) = ((x_i \cdot x_j) + c)^d \quad (5)$$

2.2 K-Nearest Neighbor (KNN)

Algoritma k-NN adalah suatu metode yang menggunakan algoritma *supervised learning* (Han, 2012). Algoritma k-NN merupakan salah satu teknik *lazy learning*. k-NN dilakukan dengan mencari kelompok k objek dalam data *training* yang paling mirip dengan objek pada data baru atau data *testing* (Leidiyana, 2013).

Secara umum untuk mendefinisikan jarak antara dua objek x dan y , digunakan rumus jarak *Euclidean* (Han & Kamber, 2012) pada persamaan 1.

$$d(x_i, x_j) = \sqrt{\sum_{i=1}^n (a_r(x_i) - y_i(x_{ij}))^2} \quad (6)$$

dengan :

$d(x_i, x_j)$ adalah Jarak *Euclidean*, x_i merupakan *record* ke- i , x_j adalah *record* ke- j , a_r adalah data ke- r , dan i, j menunjukkan data ke-1,2,..., n .

2.3 Synthetic Minority Over-Sampling Technique (SMOTE)

SMOTE merupakan metode yang diusulkan sebagai salah satu solusi dalam menangani data tidak seimbang dengan prinsip yang berbeda dengan metode *oversampling* yang telah diusulkan sebelumnya (Chawla, et. al., 2002). Jika metode *oversampling* memiliki prinsip untuk memperbanyak pengamatan secara acak, maka metode SMOTE menambah jumlah data kelas minor agar setara dengan kelas mayor dengan cara membangkitkan data buatan atau sintesis. Data sintesis tersebut dibuat berdasarkan k -tetangga terdekat (*k-nearest neighbor*). Jumlah k ditentukan dengan mempertimbangkan kemudahan dalam melaksanakannya. Pembangkitan data berskala numerik berbeda dengan kategori. Data numerik diukur jarak kedekatannya dengan *Euclidean distance* sedangkan data kategori lebih sederhana yaitu dengan nilai modus. Perhitungan jarak antar contoh kelas minor yang peubahnya berskala kategorik dilakukan dengan rumus *Value Difference Matrix* (VDM) yaitu (Barro, 2013) :

$$\Delta(X, Y) = w_X w_Y \sum_{i=1}^N \delta(x_i, y_i)^r \quad (7)$$

dengan :

$\Delta(X, Y)$ adalah jarak antara amatan X dengan Y , $w_X w_Y$ adalah bobot amatan (dapat diabaikan), N adalah banyaknya peubah sejenis. Sedangkan r bernilai 1 (jarak Manhattan) atau 2 (jarak Euclidean), $\delta(x_i, y_i)^r$ adalah jarak antar kategori, dengan rumus:

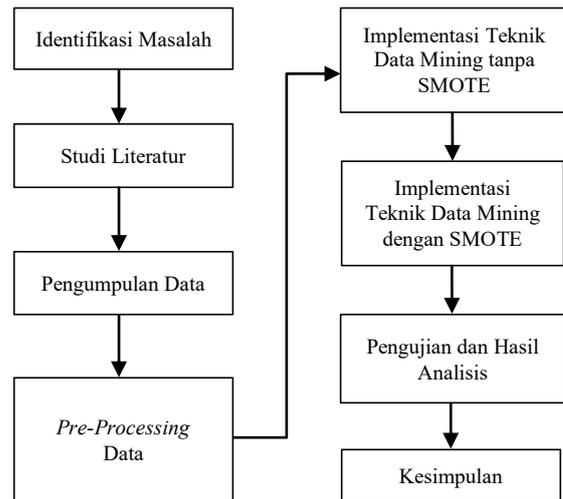
$$\delta(V_1, V_2) = \sum_{i=1}^N \left| \frac{c_{1i}}{c_1} - \frac{c_{2i}}{c_2} \right|^k \quad (8)$$

dengan $\delta(V_1, V_2)$ adalah jarak antara nilai V_1 dan V_2 , c_{1i} adalah banyaknya V_1 yang termasuk kelas i , c_{2i} adalah banyaknya V_2 yang termasuk kelas i , I adalah

banyaknya kelas; $i = 1, 2, \dots, m$. Sedangkan C_1 menotasikan banyaknya nilai 1 terjadi, C_2 menyatakan banyaknya nilai 2 terjadi, N adalah banyaknya kategori, dan K adalah konstanta yang pada umumnya bernilai 1.

3. METODE PENELITIAN

Tahapan yang dilakukan untuk menyelesaikan penelitian ini ditunjukkan oleh Gambar 1.



Gambar 1. Tahapan penelitian

Tahapan penelitian dimulai dari proses identifikasi masalah yang dilakukan dengan teknik wawancara kepada sekretaris Jurusan Sistem Informasi dan Teknik Informatika di Fakultas Ilmu Komputer Universitas Brawijaya. Selanjutnya adalah melakukan studi literatur terkait dengan metode *data mining* yang digunakan yaitu SVM, k-NN, dan SMOTE. Tahap berikutnya adalah pengumpulan data. Data yang dikumpulkan adalah data Indeks Prestasi (IP) dan Kartu Hasil Studi (KHS) mahasiswa yang berasal dari tahun akademik 2014 semester ganjil sampai tahun akademik 2016 semester genap.

Pre-processing data dilakukan dalam 2 tahap. Tahap pertama yaitu integrasi data dan tahap kedua yaitu transformasi data. Integrasi data dilakukan untuk menggabungkan data dari beberapa tabel yang berbeda. Tabel-tabel data tersebut yaitu tabel KHS dan IP untuk mengambil atribut yang diperlukan yaitu Nilai, IP, IPK, SKS, SKSK dan Semester berdasarkan mata kuliah prasyaratnya. Transformasi data dilakukan untuk mengubah nilai atribut tertentu yang sebelumnya berupa huruf menjadi angka dengan kategori tertentu. Transformasi data yang dilakukan pada atribut Nilai dan Kelas sebagaimana ditunjukkan pada Tabel 1 dan 2.

Data yang digunakan dalam penelitian ini adalah data tahun 2014 semester ganjil sampai 2015 semester genap. Hasil dari proses *pre-processing* adalah *dataset* yang terdiri dari 7 mata kuliah dengan komposisi sebagaimana ditunjukkan oleh Tabel 3.

Tahapan selanjutnya adalah implementasi teknik *data mining*. Implementasi dilakukan menggunakan software *Orange*, dengan *dataset* yang digunakan untuk proses *training* adalah data tahun 2014 semester ganjil sampai tahun 2015 semester genap. Untuk pengujian menggunakan teknik *Cross Validation*. Analisis hasil klasifikasi akan dilihat berdasarkan nilai *Area Under Curve* (AUC), *Classification Accuracy* (CA), *F₁*, *Precision*, dan *Recall*-nya.

Tabel 1. Transformasi atribut Nilai

Nilai	Nilai Pre-Process
A	1
B+	2
B	3
C+	4
C	5
D+	6
D	7
E	8

Tabel 2. Transformasi atribut Kelas

Nilai	Nilai Pre-Process
Ya	1
Tidak	0

Tabel 3. Komposisi data hasil *pre-processing*

Mata Kuliah	Jumlah kelas ya	Jumlah kelas tidak	Jumlah data
Manajemen Hubungan Pelanggan	234	726	960
Basis Data Terdistribusi	165	3656	3821
Sistem Pendukung Keputusan	189	3528	3717
Sistem Pakar	838	8286	9124
Logika Fuzzy	304	4614	4918
Arsitektur Jaringan Terkini	221	5541	5762
Jaringan Nirkabel	231	4884	5115

Proses implementasi data mining yang dilakukan adalah dengan metode SVM dan *k*-NN tanpa menggunakan teknik SMOTE. Adapun pengaturan yang dilakukan pada masing-masing metode adalah sebagai berikut. Pada implementasi *k*-NN pengaturan yang dilakukan adalah jumlah *k* sebanyak 5, dengan *metric* = *Euclidean* dan *weight* = *uniform*. Sedangkan untuk SVM pengaturan yang dilakukan adalah mengatur nilai *Cost* (C) sebesar 1,00 dan *Regression loss epsilon* (ϵ) sebesar 0,10. Untuk kernel diatur ke RBF, *numerical tolerance* sebesar 0,0010 dan *iteration limit* sebesar 100.

Selanjutnya adalah proses *mining* yang dilakukan dengan membandingkan hasil jika tidak menggunakan teknik SMOTE dan dengan menggunakan teknik SMOTE untuk mengetahui perbedaan penerapan SMOTE pada *imbalance dataset*.

1. Implementasi tanpa SMOTE

Untuk tahap pertama dilakukan proses implementasi tanpa menggunakan teknik SMOTE. *Dataset* yang telah melalui *pre-processing* akan melalui proses learning dengan menggunakan metode SVM dan *k*-NN.

2. Implementasi dengan SMOTE

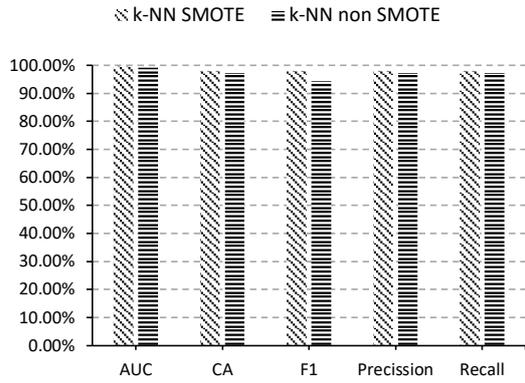
Setelah proses implementasi tanpa SMOTE dilakukan maka selanjutnya adalah melakukan proses implementasi dengan data yang telah disetarakan kelasnya melalui teknik SMOTE. Data yang telah dipre-process dan menggunakan teknik SMOTE selanjutnya akan melalui proses *learning* dengan metode yang sama.

4. HASIL DAN PEMBAHASAN

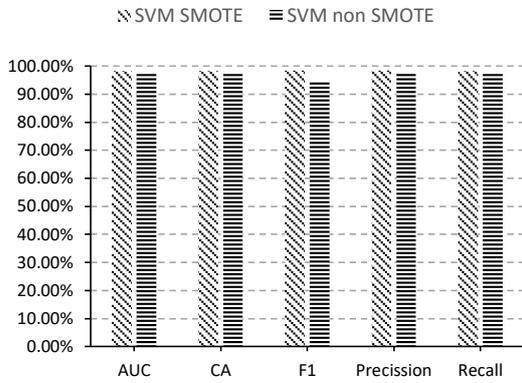
Berdasarkan proses sebelumnya yaitu penerapan metode klasifikasi SVM dan *k*-NN tanpa menggunakan SMOTE dan dengan menggunakan SMOTE dilakukan proses analisis hasil berdasarkan nilai AUC, CA, *F₁*, *Precision* dan *Recall* untuk mengukur kinerja dari klasifikasi. Analisis juga dilakukan untuk mengetahui perbedaan pengaruh penggunaan SMOTE terhadap kasus *imbalance dataset* pada metode SVM dan *k*-NN dalam mengklasifikasikan mahasiswa pengambil mata kuliah. Analisis hasil klasifikasi menggunakan 3 mata kuliah sampel dari 7 mata kuliah yang digunakan dalam proses *mining* yaitu Manajemen Hubungan Pelanggan, Sistem Pendukung Keputusan dan Jaringan Nirkabel. Adapun hasil pengujian kinerja dari ketiga mata kuliah ditunjukkan pada Gambar 2 – 7. Sedangkan hasil rata-rata nilai AUC, CA, *F₁*, *Precision* dan *Recall* dari hasil klasifikasi ditunjukkan oleh Tabel 4.

Dari hasil pengujian menggunakan teknik *Cross Validation*, menggunakan grafik hasil pengukuran kinerja klasifikasi 3 mata kuliah sampel, dapat dilihat bahwa dari ketiga mata kuliah yang dijadikan sampel menunjukkan pengaruh SMOTE terhadap nilai AUC, CA, *F₁*, *Precision*, dan *Recall*. Data yang diproses menggunakan teknik SMOTE memiliki nilai yang lebih baik dan menunjukkan kenaikan yang cukup signifikan dibandingkan tanpa menggunakan SMOTE. Untuk mata kuliah Manajemen Hubungan Pelanggan, penggunaan SMOTE memiliki kenaikan dalam semua aspek penilaian untuk setiap metode yang digunakan meskipun kenaikan dari setiap aspek tidak signifikan. Pada mata kuliah Sistem Pendukung Keputusan untuk metode *k*-NN penggunaan SMOTE menunjukkan kenaikan pada setiap aspek penilaian sementara pada metode SVM kenaikan hanya pada penilaian AUC dan *F₁*-nya saja. Sedangkan pada mata kuliah Jaringan Nirkabel penggunaan SMOTE juga menunjukkan perubahan pada setiap aspek penilaian kinerja sistem untuk metode *k*-NN. Sedangkan penggunaan SMOTE pada SVM menunjukkan penurunan nilai akurasi.

1. Manajemen Hubungan Pelanggan

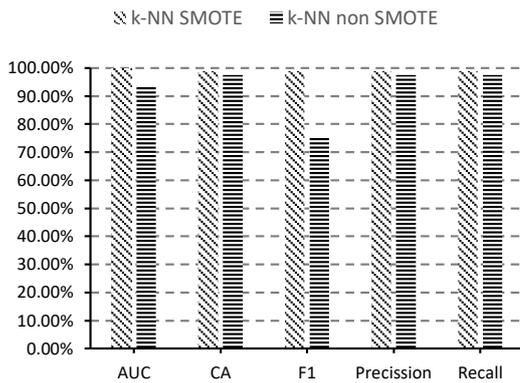


Gambar 2. Hasil Kinerja *k*-NN pada Mata Kuliah Manajemen Hubungan Pelanggan

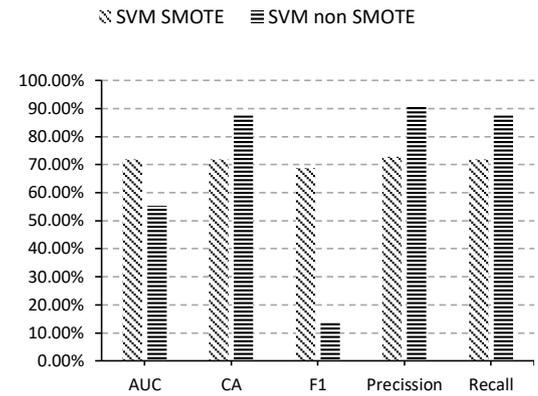


Gambar 3. Hasil Kinerja SVM pada Mata Kuliah Manajemen Hubungan Pelanggan

2. Sistem Pendukung Keputusan

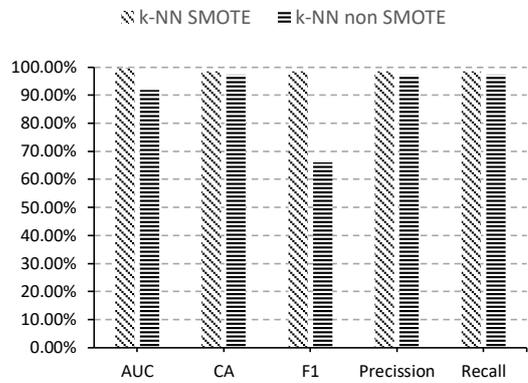


Gambar 4. Hasil Kinerja *k*-NN pada Mata Kuliah Sistem Pendukung Keputusan

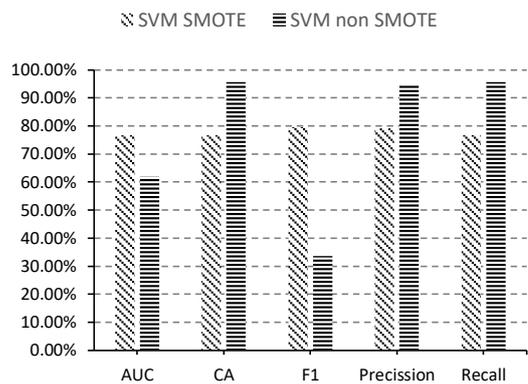


Gambar 5. Hasil Kinerja SVM pada Mata Kuliah Sistem Pendukung Keputusan

3. Jaringan Nirkabel



Gambar 6. Hasil Kinerja *k*-NN pada Mata Kuliah Jaringan Nirkabel



Gambar 7. Hasil Kinerja SVM pada Mata Kuliah Jaringan Nirkabel

Tabel 4. Hasil Pengujian Kinerja Rata-Rata Mata Kuliah

Metode	SMOTE	AUC	CA	F1	Precision	Recall
<i>k</i> -NN	Ya	99.44%	97.96%	0.9797	97.99%	97.96%
	Tidak	88.33%	96.90%	0.6016	96.34%	96.90%
SVM	Ya	78.40%	78.40%	0.7574	79.70%	78.40%
	Tidak	70.70%	89.97%	0.4283	93.76%	89.97%

Sedangkan dari Tabel 4 rata-rata nilai AUC, CA, F1, *Precision* dan *Recall*, dapat disimpulkan bahwa metode *k*-NN menunjukkan hasil yang lebih baik daripada metode SVM dalam mengklasifikasikan mahasiswa pengambil mata kuliah. Metode *k*-NN memiliki nilai yang lebih baik untuk proses klasifikasi tanpa SMOTE yaitu nilai AUC sebesar 88,33%, nilai CA sebesar 96,90%, nilai F1 sebesar 0.6016, nilai *precision* sebesar 96,34%, dan nilai *recall* sebesar 96,90%. Sementara untuk penggunaan SMOTE metode *k*-NN juga menunjukkan hasil yang lebih baik daripada metode SVM yaitu nilai AUC sebesar 99,44%, CA sebesar 97,96%, F1 sebesar 97,97, *precision* sebesar 97,99%, dan *recall* sebesar 97,96%. Akan tetapi penggunaan SMOTE memiliki penurunan nilai pada nilai CA, *precision* dan *recall* untuk metode SVM. Sehingga dapat ditarik kesimpulan bahwa metode *k*-NN memiliki kinerja yang lebih baik daripada metode SVM dalam melakukan klasifikasi mahasiswa pengambil mata kuliah baik pada data *imbalance* yang diklasifikasikan tanpa teknik SMOTE maupun data yang diproses dengan menggunakan teknik SMOTE terlebih dahulu.

5. KESIMPULAN

Penggunaan teknik SMOTE untuk menangani kasus *imbalance dataset* pada klasifikasi mahasiswa pengambil mata kuliah menunjukkan peningkatan nilai AUC, CA, F1, *precision*, dan *recall*-nya. Hal tersebut dilihat dari hasil pengujian yang dilakukan menggunakan 3 mata kuliah sampel yaitu Manajemen Hubungan Pelanggan, Sistem Pendukung Keputusan, dan Jaringan Nirkabel yang menunjukkan kecenderungan peningkatan nilai AUC, CA, F1, *precision* dan *recall*-nya. Selain itu juga terlihat dari rata-rata hasil yang menunjukkan bahwa penggunaan SMOTE mempengaruhi hasil pengujian yang dapat meningkatkan nilai AUC, CA, F1, *precision* dan *recall*.

Secara umum metode KNN memiliki kinerja yang lebih baik daripada metode SVM baik dalam penggunaan SMOTE pada *dataset* maupun tanpa penggunaan SMOTE. Hal tersebut dapat dilihat dari hasil rata-rata pengujian yang ditunjukkan oleh nilai AUC, CA, F1, *precision* dan *recall*.

6. DAFTAR PUSTAKA

- BARRO, R.A., SULVIATI, I.D., DAN AFENDI, F.M. 2013. Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu. Vol. 1(1):e9(1-6). Xplore.
- CHAWLA, V.N., BOWYER, K.W., HALL, L.O., dan KEGELMEYER, W.P. 2002. SMOTE: Synthetic Minority Over-sampling Technique. p321–357. Journal of Artificial Intelligence Research.
- CHRISTIAN, T.M., DAN AYUB, M. *Exploration of Classification Using NBTree for Predicting Students' Performance*. 2014 International Conference on Data and Software Engineering (ICODSE).
- HAN, J. DAN KAMBER, M. 2012. *Data Mining Concepts and Techniques 2nd Edition*. San Fransisco : Morgan Kaufmann Publishers.
- HOU, Y. DING, X. DAN HOU R. 2017. *Support Vector Machine Classification Prediction Model Based on Improved Chaotic Differential Evolution Algorithm*. 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery.
- LEIDIANA, H. 2013. Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. p.65-76. PIKSEL (Penelitian Ilmu Komputer Sistem Embedded dan Logic).
- OCTAVIANI, P.A., WILANDARI, Y., DAN ISPRIYANTI, D. 2014. *Penerapan Metode Klasifikasi Support Vector Machine (Svm) Pada Data Akreditasi Sekolah Dasar (Sd) di Kabupaten Magelang*. JURNAL GAUSSIAN, Volume 3, Nomor 4, Tahun 2014, Halaman 811 – 820.
- PRASETYO, E. 2014. *Data Mining, Mengolah Data Menjadi Informasi Menggunakan Matlab*. Penerbit Andi.
- PUJIANTO, U., AZIZAH, E.N., DAN DAMAYANTI, A.S. 2017. *Naive Bayes using to predict students' academic performance at faculty of literature*. 2017 5th International Conference on Electrical, Electronics and Information Engineering (ICEEIE).
- SANTOSA, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta : Graha Ilmu.
- TARUNA, S., PANDEY, M. 2014. *An Empirical Analysis of Classification Techniques for Predicting Academic Performance*. 2014 IEEE International Advance Computing Conference (IACC).
- WU, T.K., HUANG, S.C., DAN MENG, Y.R. 2008. *Evaluation of ANN and SVM classifiers as predictors to the diagnosis of students with learning disabilities*. Expert Systems with Applications 34 (2008) 1846–1856.