JINTEKS (Jurnal Informatika Teknologi dan Sains) ISSN 2686-3359 (Online)
Vol. 5 No. 3, Agustus 2023, hlm. 322 – 327

SELEKSI ATRIBUT PADA DATA TIDAK SEIMBANG NASABAH KOPERASI DENGAN OPTIMASI SMOTE DAN ADABOOST

Andreyestha[1*], Richky Faizal Amir [2], Imam Nawawi[3], Andi Taufik[4] , Eko Pramono[5],Fajar Akbar[6]
[1,3]Sistem Informasi, Universitas Bina Sarana Informatika
[2]Teknik Komputer, Universitas Bina Sarana Informatika
[5]Ilmu Komputer, Universitas Bina Sarana Informatika
[4,6]Informatika , Universitas Nusa Mandiri
email: andreyes2505@bsi.ac.id[1]

Abstrak: Prosedur kredit adalah pemberian kredit atas dasar keyakinan bank terhadap kemampuan dan kesanggupan nasabah untuk mengembalikan. Pada penelitian ini Data nasabah yang diuji dibagi menjadi 3 kelas yaitu 39 nasabah berpenghasilan RENDAH, 84 nasabah berpenghasilan SEDANG, dan 11 nasabah berpenghasilan TINGGI, dataset yang tidak seimbang tersebut dilakukan *preprocessing* menggunakan tehnik *Synthetic Minority Over-sampling (SMOTE).* Metode klasifikasi seperti *Random Forest* dan *Support Vector Machine* akan melakukan pengujian pada data nasabah koperasi. Data tersebut diuji berdasarkan atribut yang memiliki *matching value* paling tinggi menggunakan metode *Particle Swarm Optimization*, pada penelitian ini dilakukan optimasi hasil dari pengujian data nasabah menggunakan metode *Adaboost* untuk meningkatkan akurasinya. Hasil pengujian yang dilakukan dengan menggunakan metode *Random Forest* mendapat akurasi 89.05% dan algoritma *Support Vector Machine* mendapatkan akurasi 81.75%. Pada saat yang sama, pengujian dengan dua metode yang dioptimalkan dengan *Adaboost* menunjukkan peningkatan akurasi dengan hasil *Random Forest* mendapat akurasi 91.24% dan *Support Vector Machine* mendapat akurasi 82.48%. Akurasi tertinggi pada pengujian klasifikasi data nasabah koperasi didapat dari algoritma *Random Forest* yang dioptimalkan dengan *Adaboost* sebesar 91.24%.

Kata Kunci : Selekasi Atribut, *Adaboost*, *SMOTE*, Hutan Acak, Nasabah, Koperasi.

Abstract: *The credit procedure is the provision of credit on the basis of the bank's belief in the ability and ability of the customer to repay. In this study, the customer data tested was divided into 3 classes, namely 39 LOW customers, 84 MEDIUM customers, and 11 HIGH customers. Unbalanced datasets are pre-processed using the Synthetic Minority Over-sampling (SMOTE) technique. Classification methods such as Random Forest and Support Vector Machine will test cooperative customer data. The data is tested based on the attribute that has the highest matching value using the Particle Swarm Optimization method, in this research also use optimized with the Adaboost method to increase its accuracy. The results of tests carried out using the Random Forest method obtained an accuracy of 89.05% and the Support Vector Machine algorithm obtained an accuracy of 81.75%. Meanwhile, testing with two methods optimized with Adaboost showed an increase in accuracy with Random Forest getting 91.24% accuracy and Support Vector Machine getting 82.48% accuracy. The highest accuracy in the cooperative customer data classification test was obtained from the Random Forest algorithm which was optimized with Adaboost at 91.24%.*

Keywords **:** *Attribute Selection, Adaboost, SMOTE, Random Forest, Customer, Cooperative.*

PENDAHULUAN

Pemahaman mengenai koperasi adalah wadah bagi kelompok yang rentan secara ekonomi dan bersifat altruistik, memungkinkan setiap orang untuk memenuhi kewajiban keanggotaannya dan menerima kompensasi yang adil. Dari segi perbankan, pemberian pinjaman merupakan kegiatan usaha yang biasa dilakukan oleh bank umum untuk menanamkan modal bagi masyarakat. Kredit merupakan sumber pendapatan utama bank dan juga sumber risiko operasional terbesar [1]. Pengajuan kredit atau pinjaman koperasi biasanya lebih mudah dari pada bank dan tetap memenuhi syarat pokok pinjaman dengan persyaratan seperti jumlah pinjaman, jangka waktu pembayaran, jenis pekerjaan, status perkawinan, jumlah anggota keluarga, tanggungan, pendapatan, pajak, tagihan listrik, telepon, dll. Prosedur kredit adalah pemberian kredit atas dasar keyakinan bank terhadap kemampuan dan kesanggupan nasabah untuk mengembalikan. Selain itu, strategi pemasaran yang diterapkan oleh jajaran Komida belum lancar dan masih ada sebagian masyarakat yang belum mengetahuinya. Jika kredit yang diterbitkan macet, maka tindakan yang dilakukan adalah mengajukan kredit tersebut [2].

Teknik *data mining* adalah metode evaluasi data paling akurat yang dapat digunakan di gudang data. Penambangan data digunakan untuk memprediksi informasi tersembunyi menggunakan proses ekstraksi untuk meningkatkan pengambilan keputusan [3]. *Data Mining* adalah proses penggalian informasi atau pengetahuan yang berguna dari sejumlah besar data. Ada berbagai metode yang dapat digunakan untuk penambangan data. Secara umum, ada dua jenis kegiatan yang terlibat dalam proses *data mining*, yaitu prediktif dan deskriptif. [4].

Penelitian ini bertujuan untuk mengetahui nilai akurasi yang didapatkan dari beberapa metode klasifikasi seperti *Random Forest* dan *Support Vector*

 ISSN 2686-3359 (Online)

Machine yang diuji pada data nasabah koperasi. Data nasabah koperasi diuji berdasarkan atribut yang memiliki *matching value* paling tinggi menggunakan metode *Particle Swarm Optimization*, termasuk kedalam pengklasifikasian yang mempunyai nilai akurasi tinggi dengan algoritma lain yang didapat.

TINJAUAN PUSTAKA

Dalam tinjauan pustaka, akan dijelaskan literatur yang relevan dengan penelitian, serta metode-metode yang digunakan dalam penelitian.

Particle Swarm Optimization (PSO) adalah algoritma pengoptimalan pencarian grup. Gerakan setiap partikel ditentukan oleh nilai fungsi kebugaran, "arah" dan "target" gerakannya ditentukan oleh "kecepatan" setiap partikel. Kemudian, partikel-partikel tersebut beriterasi dalam ruang solusi sesuai dengan arah partikel terbaik [5].

SMOTE merupakan metode untuk menciptakan sebuah contoh dari kelas minoritas sintetis yang beroperasi di ruang fitur daripada ruang data. Dengan menduplikasi contoh kelas minoritas, *SMOTE* menghasilkan contoh sintetis baru dengan melakukan ekstrapolasi sampel minoritas yang ada dengan sampel acak yang diperoleh dari nilai k tetangga terdekat [6]. Teknik ini membantu mengatasi keterbatasan yang disebabkan oleh *oversampling* sederhana dan meningkatkan kumpulan data asli dengan cara yang secara signifikan meningkatkan pembelajaran. Oleh karena itu, teknik ini telah menunjukkan banyak manfaat yang menjanjikan. *SMOTE* dapat meningkatkan akurasi dari metode klasifikasi [7].

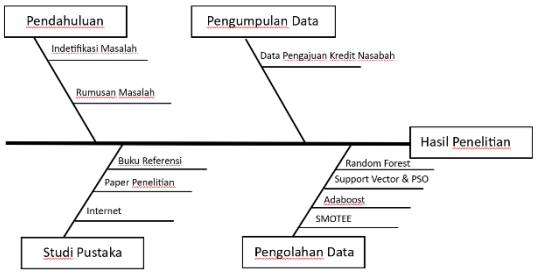
Adaboost dapat dikombinasikan dengan pengklasifikasi algoritma lain untuk meningkatkan kinerja. Menggabungkan beberapa model membantu secara intuitif ketika model berbeda satu sama lain [8]. Pohon-pohon ini sering disebut sebagai "tunggul pohon". Dengan memberikan bobot yang sama pada setiap titik data, metode ini membangun sebuah model. Skor yang salah klasifikasi kemudian diberi bobot lebih. Semua skor dengan bobot lebih tinggi lebih penting dalam sampel berikut [9].

Random Forest adalah algoritma yang umum digunakan untuk mengukur kepentingan fitur. Itu tidak hanya dapat melakukan pemilihan fitur, tetapi juga memiliki akurasi dan daya tahan klasifikasi yang baik. *Random Forest* memiliski kecepatan latihannya yang cepat, akurasi tinggi, dan tidak perlu penyetelan parameter yang rumit [10]. Algoritma *Random Forest* menggunakan pohon keputusan yang masing-masing memprediksi nilai untuk variabel yang diteliti, yang dalam penelitian ini adalah ukuran geofisika (VWC atau EC). Estimasi terbaik untuk pengukuran geofisika adalah rata-rata estimasi dari semua pohon keputusan [11].

Support Vector Machine adalah metode klasifikasi yang dikembangkan oleh Vladimir Vapnik yang menggunakan metode pembelajaran mesin *(supervised learning)* yang memprediksi kelas berdasarkan pola dari hasil proses pelatihan [12]. SVM memiliki keunggulan dibandingkan metode pembelajaran mesin lainnya karena metode ini dapat menerapkan diferensiasi linier ke data input nonlinier berdimensi tinggi, dan mencapainya dengan menggunakan fungsi kernel yang diperlukan [13].

METODE

Dalam penelitian ini penlusi melakukan beberapa tahapan penelitian yang digambarkan dengan bagan yang ada pada gambar. 1.



Gambar 1. Kerangka Peneitian

Tahap Studi Literatur ini mencari ide dari berbagai sumber seperti buku, majalah dan sumber lainnya mengkaji baik konsep maupun teori untuk menemukan sumber yang tepat dan relevan.

Permasalahan data mining terkait peminjaman nasabah koperasi diidentifikasi sehingga dapat ditentukan rencana kerja dan memutuskan data apa saja yang diperlukan dalam penelitian.

Tahap pengumpulan data dilakukan untuk mencari informasi tentang informasi yang berkaitan dengan klien atau nasabah koperasi. Atribut data yang digunakan merupakan jaminan, pekerjaan, status, jumlah tanggungan, jenis kelamin, usia, tempat tinggal, pendapatan, dan hasil survei, di antara banyak atribut data nasabah. Data yang akan digunakan adalah 150 data nasabah koperasi yang mana data cleaning atau seleksi data dilakukan untuk memfilter data *noise* untuk menemukan data yang cocok untuk data training. Setelah proses pembersihan atau pemilihan data, diperoleh 134 data bebas *noise* dan digunakan untuk data pelatihan. Kemudian 134 data uji yang digunakan berasal dari data pelatihan, dengan data dibersihkan agar bebas *noise*. Informasi ini digunakan untuk menghitung akurasinya menggunakan algoritma *Random Forest* dan *Support Vector Machine* (SVM).

Data yang terkumpul dilakukan pra-pemrosesan yang membersihkan data pada tahap ini, atau membersihkannya dari data yang tidak konsisten atau mengandung *noise*. Data nasabah dengan atribut yang tidak lengkap akan dipisahkan. Selain itu, beberapa hal dilakukan dalam proses pembersihan data.

Pengelompokan jumlah tanggungan yang dimiliki nasabah menjadi 3 kategori, yaitu ringan (1-2), sedang (3-5), banyak (6-<6). Pengelompokan usia

 ISSN 2686-3359 (Online)

nasabah menjadi 3 kategori yaitu muda (18-28 tahun), sedang (>28-40 tahun), tua (>40 tahun). Pengelompokan penghasilan yang dimiliki nasabah menjadi 3 kategori yaitu rendah (100 rb – 500 rb), sedang (1 jt – 5 jt), tinggi (>5 jt). Pengelompokan pinjaman nasabah yang berupa angka menjadi 4 kategori yaitu rendah (100 rb- 500 rb), sedang (600 rb – 1jt), tinggi (1,5 jt – 5 jt), dan sangat tinggi (6 jt – 10 jt). Klasifikasi Pekerjaan menjadi 2 kategori yaitu BUMN dan Non BUMN. Kemudian hasil dari cleaning data ini disusun dalam bentuk format csv pada tools Notepad++.

Transformasi data merupakan konversi dan integrasi data yang digunakan dalam proses penambangan. Pada titik ini, format data diubah menjadi format yang sesuai dengan metode penambangan data yang digunakan. WEKA mendukung beberapa format file input, termasuk format CSV. CSV adalah file teks berformat khusus yang dapat menyimpan data dalam tabel terstruktur.

Saat mencari solusi, beberapa atribut digunakan untuk klasifikasi. Namun dari segi kinerja, ada kemungkinan atribut yang tidak memiliki matching value dalam data mining dapat menimbulkan kebingungan dalam proses *data mining* jika atribut tersebut dimasukkan dalam proses klasifikasi. Oleh karena itu, atribut harus dipilih sedemikian rupa sehingga atribut dengan nilai yang tidak sesuai atau terkait dapat diidentifikasi. Pada tahap ini metode yang digunakan yaitu *Particle Swarm Optimization* (PSO).

Pada fase *Preprocessing*, Weka digunakan sebagai *tool* pendukung, sebelum algoritma klasifikasi *Random Forest* dan *Support Vector Machine* (SVM) melakukan prosedur klasifikasi, diperlukan *preprocessing* untuk menyeimbangkan dataset menggunakan metode *Synthetic Minority Over-sampling (SMOTE)* berdasarkan class Penghasilan yang berisi kriteria Rendah, Sedang, dan Tinggi. Penyeimbangan data ini dilakukan agar hasil akurasi menjadi lebih baik.

Proses Klasifikasi dilakukan pada data dengan algoritma *Random Forest* dan algoritma *Support Vector Machine* (SVM) yang dioptimasi dengan metode *Adaptive Boosting* (*Adaboost*) menggunakan tools Weka yang diperoleh dari data training dan test.

Model diuji menggunkana *tool* Weka, hasil yang diperoleh dengan metode *Random Forest* dan *Support Vector Machine* (SVM) diuji menggunakan *10-K-Fold Cross Validation*. Validasi silang adalah bentuk sederhana dari rekayasa statistik. Jumlah lipatan standar yang digunakan untuk memprediksi tingkat kesalahan data menggunakan *10 K-Fold Cross Validation*.

Hasil analisis dievaluasi keakuratan data model yang diperoleh dengan dihitung menggunakan metode *Random Forest* dan *Support Vector Machine* (SVM) menggunakan metode 10-fold cross-validation dengan optimasi *Adaboost.*

Input : : Biarkan D menjadi dataset yang disertakan

$\{(a1, b1), (a2, b2), \ldots\ldots (am, bm)\};$

Biarkan λ menjadi algoritma pembelajaran (basis).
Biarkan T menjadi Jumlah total putaran pembelajaran.
Process:

$D1(i) = 1/m$
$for\ time\ =\ 1, \ldots, T;$
$ht\ =\ \lambda\ (D, Dt);$

klasifikasi yang lemah dilatih dengan Distribusi Dt

$\in t\ =\ PrPri{\sim}Dt[\ \not{h}t(ai \neq bi)];$

Ukuran kesalahan (entropi)

$\propto t\ =\ ½\ln(1-\in t \notin t)$; % tentukan berat ht

$Dt + 1(i)\ =\ Dt(i)Zt\ *\{exp(-\propto t)\ if\ ht(ai)$
$= biexp(\propto t)\ if\ ht(ai) \neq bi$
$=\ Dt(i)exp(-\propto tyht(ai)/Zt$

Outcome:

$H(a) =\ \text{sign}\left(\sum Tt = 1 \propto tht(b)\right)$ [9]

Dengan teknik ini, kumpulan data dibagi secara acak menjadi K partisi. Kemudian dilakukan percobaan K, pada setiap percobaan data dari bagian Kth digunakan sebagai data uji dan bagian yang tersisa sebagai data pelatihan. Jika hasil maksimal diambil. Selain itu, format tampilan setiap atribut dapat dilihat pada alat WEKA.

HASIL DAN PEMBAHASAN

Analisis masalah bertujuan untuk mengidentifikasi dan mengevaluasi sistem yang sedang berjalan berdasarkan kebutuhan koperasi guna mencari solusi atas permasalahan yang dihadapi koperasi. Metode *Support Vector Machine* (SVM) dan *Random Forest* digunakan dalam penelitian ini. Analisis dilakukan untuk menemukan informasi baru dari data yang ada. Saat menghitung dengan *Support Vector Machine* (SVM) dan *Random Forest*, data yang diperoleh dibagi menjadi beberapa kelas.

Data yang diperoleh berdasarkan observasi yang dilakukan meliputi data nasabah sebanyak 150. Data yang digunakan adalah data pelanggan kolaboratif tahun 2019 dan memiliki atribut. garansi BPKP sepeda motor, sertifikat properti, sertifikat bangunan, KK (pekerjaan, jumlah anak, ID (identitas, jenis kelamin, usia, tempat tinggal), pendapatan, hasil survei, dan pinjaman, data nasabah disajikan pada Tabel 1.

 ISSN 2686-3359 (Online)

Tabel 1. Data Nasabah

No	Nama Karyawan	Jk	Usia	Penghasilan	Jaminan	Tanggungan	Hasil survei	Pinjaman
1	ACHMAD SAUKI	L	27	3.850.000	SERTIFIKAT	3	BAIK	RENDAH
2	A. RUSMAN	L				5		
3	ABDUL RAZAK	L	49	5.150.500	MOTOR	6	CUKUP BAIK	SANGAT TINGGI
4	ABDI PRANATA	L	41	5.325.000	MOTOR	4	BAIK	TINGGI
5	A. KADIR MUNIF	L	35	4.250.750	MOTOR	3	BAIK	SEDANG

Seleksi atribut dilakukan untuk mencari kemungkinan atribut yang tidak memiliki matching value atau sebaliknya dalam proses klasifikasi, pada tahap ini metode yang digunakan yaitu *Particle Swarm Optimization* (PSO). Metode ini mengoptimalkan masalah dengan secara iteratif mencoba meningkatkan solusi kandidat sehubungan dengan ukuran kualitas yang diberikan sesuai dengan kriterisa yang ditentukan yaitu, Jenis kelamin, penghasilan, jaminan tanaggungan dan hasil survey [14].

number of folds (%) attribute
0(0 %) 1 NO
6(60 %) 2 Jk
0(0 %) 3 Usia
0(0 %) 4 Status
0(0 %) 5 Domisili
0(0 %) 6 Pekerjaan
10(100 %) 7 Penghasilan
10(100 %) 8 Jaminan
10(100 %) 9 Tanggungan
2(20 %) 10 Hasil survey

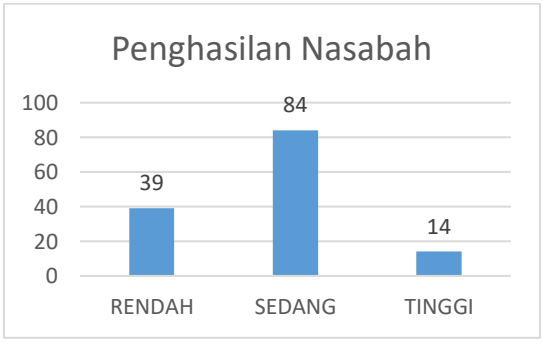
Dari atribut yang diuji, Penghasilan, Jaminan, dan Tanggungan memiliki tingkat kecocokan yang paling tinggi, pada penelitian ini Penghasilan akan diuji dalam klasifikasi untuk mendapatkan akurasi terbaik seperti hasil yseleksi yang ditampilkan pada table.2.Data Seleksi.

Tabel 2. Data Seleksi

No	Nama Karyawan	Jk	Usia	Penghasilan	Jaminan	Tanggungan	Hasil survei	Pinjaman
1	ACHMAD SAUKI	L	27	3.850.000	SERTIFIKAT	3	BAIK	RENDAH
2	A. RUSMAN	L	-	-	-	5	-	-
3	ABDUL RAZAK	L	49	5.150.500	MOTOR	6	CUKUP BAIK	SANGAT TINGGI
4	ABDI PRANATA	L	41	5.325.000	MOTOR	4	BAIK	TINGGI
5	A. KADIR MUNIF	L	35	4.250.750	MOTOR	3	BAIK	SEDANG

Preprocessing merupakan salah satu langkah yang penting dalam proses data mining. Tahapan yang dilakukan adalah pembersihan data, integrasi data, pemilihan data, dan transformasi data. Untuk data yang tidak lengkap yaitu data atribut, hapus data yang terdapat *noise*, bersihkan data, edit atribut untuk integrasi data, hapus atribut yang tidak digunakan untuk analisis, filter data, dan ganti data konversi format data menurut teknik data mining. Pada tahapan ini juga dilakukan pengolahan data dengan menyeimbangkan data yang tidak seimbang menggunakan metode *SMOTE*[15].

Data nasabah dilakukan *cleaning* data atau penyeleksian data untuk mengetahui data yang bisa digunakan untuk data training dengan menghilangkan data yang mengalami *noise*, misalnya redudansi data dan data inkonsisten. Pada Tabel 2 terdapat data yang digarisbawahi dengan warna merah, menunjukkan contoh data yang tidak lengkap atau terdapat *noise* yang mencegah data tersebut digunakan sebagai data pelatihan. Setelah pembersihan data atau penyaringan data, diperoleh 134 data bebas *noise*. Oleh karena itu, dalam penelitian ini data yang digunakan hanya 134 data nasabah sebagai data pelatihan.



Gambar 2. Data Penghasilan Nasabah Setelah Preprocessing

Data nasabah dibagi menjadi 3 kelas yaitu 39 nasabah berpenghasilan RENDAH, 84 nasabah berpenghasilan SEDANG, dan 11 nasabah berpenghasilan TINGGI, dataset yang tidak seimbang tersebut dilakukan preprocessing menggunakan tehnik *Synthetic Minority Over-sampling (SMOTE)* sehingga data disesuaikan sesuai dengan 3 kelas tersebut terlihat pada gambar IV.1. Atribut yang akan digunakan pada seluruh data nasabah yaitu jenis kelamin, usia, status, domisili, pekerjaan, penghasilan, jaminan, tanggungan, hasil survei, dan pinjaman.

Transformasi data dilakukan untuk mengubah format data menjadi format yang nantinya sesuai dengan teknik data mining *Random Forest*, dan metode *Support Vector Machine* (SVM).

Tabel 3. Transformasi Data

Data Awal	Transformasi	Keterangan
Usia 18-28 Tahun	Muda	Usia 18 sampai 28 Tahun termasuk muda
Usia >28-40 Tahun	Sedang	Usia diatas 28 sampai 40 Tahun termasuk sedang

 ISSN 2686-3359 (Online)

Usia >40 Tahun	Tua	Usia diatas 40 Tahun termasuk tua
Pekerjaan Pegawai	BUMN	Pekerjaan sebagai pegawai termasuk BUMN
Pekerjaan Cs Bandara, *Airline, Outsourcing* dll	Non BUMN	Pekerjaan sebagai Cs Bandara, *Airline,Outsourcing* dll termasuk Non BUMN
Penghasilan 100 rb-500 rb	Rendah	Penghasilan 100 rb sampai 500 rb termasuk Rendah
Penghasilan 1 jt- 5 jt	Sedang	Penghasilan 1 jt sampai 5 jt termasuk Sedang
Penghasilan > 5 jt	Tinggi	Penghasilan diatas 5 jt termasuk Tinggi
Tanggungan 1-2	Ringan	Tanggungan 1 sampai 2 termasuk Ringan
Tanggungan 3-5	Sedang	Tanggungan 3 sampai 5 termasuk Sedang
Tanggungan 6-<6	Tinggi	Tanggungan 6 sampai diatas 6 termasuk Tinggi
Pinjaman 100 rb-500 rb	Rendah	Penghasilan 100 rb sampai 500 rb termasuk Rendah
Pinjaman 600 rb - 1 jt	Sedang	Penghasilan 600 rb sampai 1 jt termasuk Sedang
Pinjaman 1.5 jt -5 jt	Tinggi	Penghasilan 1 jt sampai 5 jt termasuk Tinggi
Pinjaman 6 jt-10 jt	Sangat Tinggi	Penghasilan 1.5 jt sampai 10 jt termasuk Sangat Tinggi

Data pelatihan dan data uji digunakan sebagai proses penambangan data. Data pelatihan digunakan sebagai data model untuk membangun model penambangan. Data yang digunakan perlu diberi label terlebih dahulu dan juga dilakukan *pre-processing* sedangkan data uji digunakan untuk pengujian dalam penerapan model klasifikasi. Data yang digunakan tidak perlu diberi tanda, tetapi harus melalui pre-processing. Pada Tabel 4, data pelatihan yang digunakan untuk menghitung akurasi dari metode *Random Forest* dan *SVM*.

Tabel 4. Data Training

NO	Jk	Usia	Penghasilan	Jaminan	Tanggungan	Hasil survei	Pinjaman
1	L	SEDANG	SEDANG	MOTOR	SEDANG	BAIK	SEDANG
2	L	TUA	TINGGI	MOTOR	SEDANG	BAIK	TINGGI
3	L	TUA	TINGGI	MOTOR	BANYAK	CUKUP BAIK	SANGAT TINGGI
4	L	MUDA	SEDANG	SERTIFIKAT	SEDANG	BAIK	RENDAH
5	L	SEDANG	SEDANG	MOTOR	SEDANG	BAIK	SEDANG

Pengklasifikasian dengan algoritma *Random Forest* dilakukan dengan *cross-validation* 10 poin. Berikut adalah hasil klasifikasi menggunakan WEKA *tool*:

Tabel 5. Hasil Klasifikasi *Random Forest*

Perhitungan	Persentase
Akurasi	89.05 %
Sensitivitas/Recall/Tprate	0,891
Fprate	0,85
Precision/PPV	0,896
F-Measure	0,892
Area under ROC/ AUC	0,947

Hasil klasifikasi algoritma *Random Forest* menggunakan *10 cross-validation* menunjukkan bahwa hasil prediksi benar dengan akurasi sebesar 89.05 % dan prediksi salah dengan persentasi 10.94 %.

Klasifikasi menggunakan algoritma *Support Vector Machine*, pada pengujian ini kernel yang digunakan adalah *type polynomial*. pengujian juga dilakukan dengan menggunakan *10-cross validation*. Berikut merupakan hasil klasifikasi menggunakan tools WEKA:

Tabel 6. Hasil Klasifikasi Algoritma *SVM*

Perhitungan	Persentase
Akurasi	81.75 %
Sensitivitas/Recall/Tprate	0,818
Fprate	0,207
Precision/PPV	0,816
F-Measure	0,814
Area under ROC/ AUC	0,805

Hasil klasifikasi Algoritma *Support Vector Machine* pada Tools WEKA dengan menggunakan *10-cross validation* menunjukan hasil prediksi benar dengan akurasi sebesar 81.75 % dan prediksi salah dengan persentasi 18.24 %.

Klasifikasi menggunakan Algoritma *Support Vector Machine* dengan optimalisasi *Adaboost* dilakukan dengan menggunakan *10-cross validation*. Berikut merupakan hasil klasifikasi menggunakan *tool* WEKA:

Tabel.7. Hasil Klasifikasi SVM dengan *Adaboost*

Perhitungan	Persentase
Akurasi	82.48 %
Sensitivitas/Recall/Tprate	0,825
Fprate	0,167
Precision/PPV	0,823
F-Measure	0,824
Area under ROC/ AUC	0,926

Hasil klasifikasi Algoritma *Support Vector Machine* dengan optimalisasi *Adaboost* menggunakan *10-cross validation* menunjukkan peningkatan dibandingkan dengan hasil menggunakan algoritma *Support Vector Machine* sebelumnya, dimana akurasi meningkat sebesar 0.73%, prediksi benar dengan akurasi sebesar 82.48% dan prediksi salah dengan persentasi 17.51 %.

 ISSN 2686-3359 (Online)

Klasifikasi dengan algoritma *Random Forest* dengan optimasi *Adaboost* dilakukan dengan menggunakan *10-cross validation*. Berikut adalah hasil klasifikasi dengan menggunakan WEKA tool:

Tabel 8. Hasil Klasifikasi *Random Forest* dengan *Adaboost*

Perhitungan	Persentase
Akurasi	91.24 %
Sensitivitas/Recall/Tprate	0,912
Fprate	0,067
Precision/PPV	0,916
F-Measure	0,913
Area under ROC/ AUC	0,954

Hasil klasifikasi algoritma *Random Forest* dengan optimasi *Adaboost* menggunakan *10-cross validation* menunjukan peningkatan dari hasil penggunakan algoritma tersebut, dimana akurasi meningkat sebesar 2.19%, prediksi benar dengan akurasi sebesar 91.24% dan prediksi salah dengan persentasi 8.75 %.

KESIMPULAN DAN SARAN

Penelitian ini dapat menemukan metode dengan akurasi terbaik pada algoritma *Random Forest* dan *Support Vector Machine* untuk mendapatkan hasil yang paling akurat dari informasi yang diberikan oleh nasabah koperasi dalam melakukan peminjaman. Data nasabah koperasi ini banyak mengalami *noise*, hal ini diketahui melalui proses data cleaning terdapat 134 data nasabah yang dapat digunakan. Data nasabah kooperasi juga tidak seimbang sehingga mempengaruhi akurasi komputasi dari metode yang diuji.

Pada hasil pengujian yang dilakukan dengan menggunakan metode *Random Forest* menggunakan *10-cross validation* mendapat akurasi 89.05% dan algoritma *Support Vector Machine* menggunakan *10-cross validation* mendapatkan akurasi 81.75%. Pada saat yang sama, pengujian dengan dua metode yang dioptimalkan dengan *Adaboost* menunjukkan peningkatan akurasi dengan hasil *Random Forest* mendapat akurasi 91.24% dan *Support Vector Machine* mendapat akurasi 82.48%. Akurasi tertinggi pada pengujian klasifikasi data nasabah koperasi didapat dari algoritma *Random Forest* yang dioptimalkan dengan *Adaboost* sebesar 91.24%.

DAFTAR PUSTAKA

[1] R. Oktapiani, D. Prayudi, A. Fajria, N. S. Z. Nufus, and R. N. Lestari, "Sistem Pendukung Keputusan Untuk Menentukan Managemen Kelayakan Pemberian Kredit Di Bank Mandiri Taspen Sukabumi Menggunakan Metode Analytic Hierarchy Process," *Indones. J. Softw. Eng.*, vol. 8, no. 1, pp. 36–45, 2022, doi: 10.31294/ijse.v8i1.12054.

[2] S. B. Riono, "Jurnal Ilmiah Manajemen Dan Kewirausahaan," *Jimak*, vol. 1, no. 3, pp. 2809–2406, 2022.

[3] M. Zhang, J. Fan, A. Sharma, and A. Kukkar, "Data mining applications in university information management system development," *J. Intell. Syst.*, vol. 31, no. 1, pp. 207–220, 2022, doi: 10.1515/jisys-2022-0006.

[4] A. H. M. Alaidi, R. M. Al_airaji, H. T. S. ALRikabi, I. A. Aljazaery, and S. H. Abbood, "Dark Web Illegal Activities Crawling and Classifying Using Data Mining Techniques," *Int. J. Interact. Mob. Technol.*, vol. 16, no. 10, pp. 122–139, 2022, doi: 10.3991/ijim.v16i10.30209.

[5] X. Wang and W. Yao, "applied sciences A Discrete Particle Swarm Optimization Algorithm for Dynamic Scheduling of Transmission Tasks," *MDPI*, vol. 13, no. 7, 2023.

[6] S. Susanti, "Klasifikasi Kemampuan Perawatan Diri Anak dengan Disabilitas Menggunakan SMOTE Berbasis Neural Network," *J. Inform.*, vol. 6, no. 2, pp. 175–184, 2019, doi: 10.31311/ji.v6i2.5798.

[7] N. Thanh-Long, Tran-Minh, and L. Hong-Chuong, "A Back Propagation Neural Network Model with the Synthetic Minority Over-Sampling Technique for Construction Company Bankruptcy Prediction," *Int. J. Sustain. Constr. Eng. Technol.*, vol. 13, no. 3, pp. 68–79, 2022, doi: 10.30880/ijscet.2022.13.03.007.

[8] J. L. Putra and A. Subekti, "Prediksi Kinerja Siswa Pada E-Learning Moodle Platform Menggunakan Algoritma Adaptive Boosting," *J. Inform.*, vol. 10, no. 1, 2023.

[9] M. T. Ramakrishna, V. K. Venkatesan, I. Izonin, M. Havryliuk, and C. R. Bhat, "Homogeneous Adaboost Ensemble Machine Learning Algorithms with Reduced Entropy on Balanced Data," *MDPI*, vol. 25, no. 2, 2023, doi: 10.3390/e25020245.

[10] H. Fei *et al.*, "Cotton Classification Method at the County Scale Based on Multi-Features and Random Forest Feature Selection Algorithm and Classifier," *MDPI*, vol. 14, no. 4, 2022, doi: 10.3390/rs14040829.

[11] Y. Guan, K. Grote, J. Schott, and K. Leverett, "Prediction of Soil Water Content and Electrical Conductivity using Random Forest Methods with UAV Multispectral and Ground-coupled Geophysical Data," *MDPI*, vol. 14, no. 4, 2022, doi: 10.3390/rs14041023.

[12] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.

[13] R. A. Saputra, D. Puspitasari, and T. Baidawi, "Deteksi Kematangan Buah Melon dengan Algoritma Support Vector Machine Berbasis Ekstraksi Fitur GLCM," *J. Infortech*, vol. 4, no. 2, 2022.

[14] M. Syarif, "Vol. XII No. 2, September 2015 Jurnal Techno Nusa Mandiri," *Techno Nusa Mandiri*, vol. XII, no. 2, pp. 21–26, 2015.

[15] A. A. Arifiyanti and E. D. Wahyuni, "Smote: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining," *SCAN - J. Teknol. Inf. dan Komun.*, vol. 15, no. 1, pp. 34–39, 2020, doi: 10.33005/scan.v15i1.1850.

SURAT TUGAS

291/B.01/LPPM-UNM/III/2023

Tentang

PENELITIAN YANG DIPUBLIKASIKAN DALAM JURNAL ILMIAH
Periode Maret - Agustus 2023

Menulis pada JINTEKS (Jurnal Informatika Teknologi dan Sains), Vol. 5, No. 3, Agustus 2023
e-ISSN 2686-3359

Judul :

Seleksi Atribut Pada Data Tidak Seimbang Nasabah Koperasi Dengan Optimasi Smote Dan Adaboost

Menimbang :

1. Bahwa perlu diadakan pelaksanaan Tridharma Perguruan Tinggi dalam bentuk Penelitian.
2. untuk Keperluan pada butir 1 (satu) diatas, maka perlu dibentuk tugas yang berkaitan dengan penelitian yang dipublikasikan dalam Jurnal Ilmiah.

MEMUTUSKAN

Pertama :

Menugaskan kepada saudara/i

Fajar Akbar S.Kom, M.Kom

Sebagai Penulis yang mempublikasikan Penelitianya pada Jurnal Ilmiah.

Kedua :

Mempunyai tugas sbb:
Melaksanakan Tugas yang diberikan dengan penuh rasa tanggung jawab.

Ketiga :

Keputusan ini berlaku sejak tanggal ditetapkan, dengan ketentuan apabila dikemudian hari terdapat kekeliruan akan diubah dan diperbaiki sebagaimana mestinya

Jakarta,1 Maret 2023
Ketua LPPM
Universitas Nusa Mandiri

Andi Saryoko, M.Kom

Tembusan

- Rektor Universitas Nusa Mandiri
- Arsip
- Ybs