

Penerapan Metode *Ensemble* dalam Penentuan Persetujuan Kredit Bank

Muhammad Nur Akbar*¹, M. Hasrul H², A. Muh. Nur Hidayat³

^{1,2,3}Program Studi Teknik Informatika Universitas Islam Negeri Alauddin Makassar
e-mail: muhammad.akbar@uin-alauddin.ac.id

Abstract

The development of information technology in the banking business sector is urgently needed to improve service quality by providing comfort, security, and convenience, one of which is the credit product available at a bank. The banking sector is facing a problem, namely that it must increase the volume of lending, but must also reduce the risk of problem loans. To minimize non-performing loans, banks are required to carry out the proper analysis. One of these problems can overcome by utilizing information technology to support decision-making, in the form of data mining applications. Ensemble Classifier with majority voting is offered as a solution with an average accuracy score of 88.75% which has been tested for imbalanced class problems. After implementing 274 customer data testing, a decision was obtained that 109 customers were declared APPROVED and 165 customers were declared REJECTED..

Keyword: Classification; Creditworthiness; Ensemble Classifier; Imbalanced Class

Abstrak

Perkembangan teknologi informasi dalam bidang usaha perbankan sangat diperlukan dalam rangka meningkatkan kualitas layanan dengan memberikan kenyamanan, keamanan, dan kemudahan salah satunya pada produk kredit yang terdapat pada suatu bank. Pihak perbankan dihadapkan pada suatu masalah yaitu harus meningkatkan volume pemberian kredit namun juga harus dapat mengurangi resiko kredit bermasalah. Untuk meminimalisir kredit bermasalah pihak bank dituntut melakukan analisis yang tepat. Hal tersebut dapat diatasi salah satunya dengan memanfaatkan teknologi informasi sebagai pendukung pengambilan keputusan, berupa aplikasi data mining. *Ensemble Classifier* dengan *majority voting* ditawarkan sebagai solusi dengan nilai rata-rata akurasi sebesar 88.75% yang telah diuji pada kasus *imbalanced class*. Setelah diimplementasikan pada data testing sebanyak 274 nasabah, diperoleh keputusan 109 nasabah dinyatakan SETUJU dan 165 nasabah dinyatakan di TOLAK.

Kata kunci: Klasifikasi; kelayakan kredit; *Ensemble Classifier*; *Imbalanced Class*

1. Pendahuluan

Bank merupakan suatu lembaga keuangan yang usaha pokoknya memberi kredit dan jasa dalam lalu lintas pembayaran dan peredaran uang [1]. Bank memiliki peran yang sangat penting dalam menggerakkan roda perekonomian nasional. Di Indonesia pembiayaan dunia usaha masih didominasi oleh penyaluran kredit pihak perbankan yang dianggap mampu mendorong pertumbuhan ekonomi [2]. Di dalam perusahaan perbankan ada beberapa produk yang ditawarkan berupa tabungan, giro, deposito, dan kredit, yang merupakan aset sebagai pendapatan perusahaan tersebut.

Kredit merupakan salah satu pendapatan utama yang diterima oleh perusahaan perbankan. Aset yang menghasilkan pemasukan pada perusahaan berupa angsuran dan bunga setiap bulannya oleh nasabah dengan perjanjian tertentu sesuai dengan tujuan dan kebutuhan. Kredit dapat membantu nasabah dalam pengembangan usaha agar lebih maju, sebagai imbalan bank akan memperoleh bunga dari pinjaman tersebut. Umumnya setiap bank memiliki produk yang sama, yang berbeda hanya dari segi layanan, sebab setiap bank memiliki standar yang berbeda-beda dalam melayani nasabahnya. Namun sebelum pemberian kredit, terlebih dahulu dilakukan analisis berupa studi kelayakan (*feasibility study*) atas perusahaan atau perorangan sebagai pemohon kredit.

Ketika sebuah kredit dinyatakan diterima, ada dua kemungkinan yang terjadi. Pertama, nasabah dapat membayar angsuran dengan lancar hingga melunasi kewajibannya. Kedua, nasabah mengalami kendala dalam pembayaran atau disering dikenal dengan istilah “kredit macet”. Pihak perbankan dihadapkan pada suatu masalah yaitu harus meningkatkan volume pemberian kredit namun juga harus dapat mengurangi resiko kredit bermasalah. Untuk meminimalisir kredit bermasalah pihak bank dituntut melakukan analisis yang tepat agar tidak memberikan pinjaman kepada nasabah yang berpotensi “kredit macet”. Analisis tersebut umumnya didasarkan pada karakter, kapasitas, modal, jaminan, serta kondisi ekonomi dari nasabah.

Beberapa bank memberikan kredit kepada nasabah hanya berdasarkan intuis. Hal ini tentu sangat beresiko terhadap kelancaran kredit. Oleh karena itu dibutuhkan sebuah metode yang akurat yang dapat membantu dalam penentuan pemberian kredit. Hal tersebut dapat diatasi salah satunya dengan memanfaatkan teknologi informasi sebagai pendukung pengambilan keputusan, berupa aplikasi data mining. Data mining adalah sebuah proses menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning dalam mengidentifikasi informasi dan pengetahuan terkait dari berbagai basis data yang besar [3]. Pengetahuan yang dimaksud di sini berupa pola tersembunyi yang belum ditemukan sebelumnya, yaitu pola karakter atau profil dari nasabah yang menerima kredit ataupun yang ditolak.

Dalam penelitian ini akan digunakan pengujian data training terhadap beberapa algoritma klasifikasi, yaitu KNN, *naive bayes*, C4.5 serta *ensemble classifier* yang merupakan gabungan dari tiga algoritma sebelumnya. Dataset yang digunakan berupa data 416 nasabah yang akan dijadikan data *training* dalam membangun pengetahuan atau pola klasifikasi. Terdapat pula data 274 nasabah yang akan ditentukan apakah permohonan kreditnya ditolak atau diterima dengan memanfaatkan pola klasifikasi yang dibangun sebelumnya.

Beberapa penelitian yang telah dilakukan dengan dataset yang sama, diantaranya menggunakan *hybrid SVM* [4] dengan nilai akurasi 88,70% namun model klasifikasi yang dihasilkan relatif sulit dipahami sehingga selanjutnya diujikan dengan algoritma C4.5 [5] dengan nilai akurasi 89,18% dan model yang relatif mudah dipahami karena berbentuk aturan. Namun dari beberapa penelitian sebelumnya peneliti tidak memperhitungkan keseimbangan distribusi kelas pada dataset. Oleh sebab itu penelitian ini berfokus pada penerapan *ensemble classifier* yang memiliki kelebihan dalam menangani kasus *imbalanced class* yang banyak ditemukan dalam domain dunia nyata (*real world problem*) [6].

Imbalanced class merupakan kondisi adanya ketidakseimbangan dalam jumlah data training antara dua kelas yang berbeda, salah satu kelasnya memiliki jumlah yang sangat besar (*majority class*) sedangkan kelas yang lainnya memiliki jumlah data yg lebih kecil atau bahkan sangat kecil (*minority class*). Tantangan pada kasus *imbalanced* yaitu sebagian besar *classifier* secara umum akan memaksimalkan akurasi prediksi, tetapi evaluasi model yang didapat melalui akurasi tidak sesuai untuk permasalahan data *imbalanced* [7]. Hal ini karena model yang didapatkan dari *classifier* dengan data *imbalanced* tanpa ada modifikasi *classifier* atau data *training* akan memprediksi hampir semua *record* sebagai kelas mayor sehingga hasil prediksi mempunyai akurasi tinggi tetapi hanya untuk kelas mayor [8].

Pada umumnya pengklasifikasian banyak berfokus pada *single classifier*, namun dengan menggunakan metode *ensemble* yaitu dengan menggabungkan beberapa *classifier* dapat meningkatkan akurasi yang dihasilkan [9]. Hal ini disebabkan karena setiap *classifier* memiliki kelebihan yang berbeda-beda, dan jika dikombinasikan diharapkan akan mencapai hasil yang lebih baik.

2. Metode Penelitian

Dataset yang digunakan merupakan data aplikasi persetujuan kredit dari suatu bank. Untuk menentukan persetujuan pengeluaran kredit, pihak bank mempertimbangkan beberapa hal yang berkaitan dengan nasabah. Dataset terdiri dari berbagai macam tipe data, beberapa atribut bertipe kontinu, dan sebagian lainnya bertipe nominal. Dalam dataset juga ditemukan beberapa *missing value*.

Data dibagi menjadi 2 macam yaitu data *training* dan data *testing*. Jumlah data *training* sebanyak 416 data, sedangkan data *testing* sebanyak 274 data. Jumlah atribut sebanyak 15 atribut dengan 1 kelas atribut. Nilai data *training* diperlihatkan pada tabel 1.

Tabel 1. Data Training

Nasabah	X1	X2	X3	.	.	X14	X15	Class
1	1	30.83	0	.	.	202	0	SETUJU
2	2	58.67	4.46	.	.	43	560	SETUJU
3	2	24.5	0.5	.	.	280	824	SETUJU
.
.
414	2	?	11.25	.	.	?	5200	TOLAK
415	1	16.5	0.125	.	.	132	0	TOLAK
416	2	27.33	1.665	.	.	340	1	TOLAK

Keterangan Tabel :

a) Atribut :

X1 : 1, 2 (nominal)

X2 : kontinu

X3 : kontinu

X4 : 1, 2, 3, 4 (nominal)

X5 : 1, 2, 3 (nominal)

X6 : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 (nominal)

X7 : 1, 2, 3, 4, 5, 6, 7, 8, 9 (nominal)

X8 : kontinu

X9 : 1, 2 (nominal)

X10 : 1, 2 (nominal)

X11 : kontinu

X12 : 1, 2 (nominal)

X13 : 1, 2, 3 (nominal)

X14 : kontinu

X15 : kontinu

X16 : class nominal (SETUJU, TOLAK)

b) Missing Value :

Persentase masing-masing atribut dari jumlah keseluruhan data

X1 = 3 (1%)

X9 = 0

X2 = 8 (2%)

X10 = 0

X3 = 0

X11 = 0

X4 = 2 (1%)

X12 = 0

X5 = 2 (1%)

X13 = 0

X6 = 2 (1%)

X14 = 7 (1%)

X7 = 2 (1%)

X15 = 0

X8 = 0

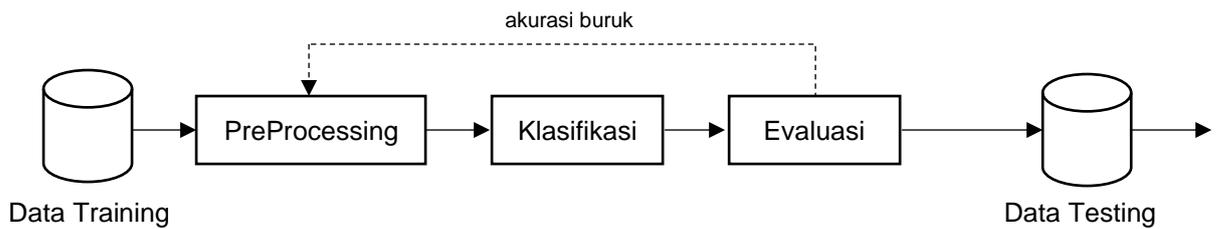
X16 = 0

Tabel 2. Data Testing (data yang classnya akan diprediksi)

Nasabah	X1	X2	X3	.	.	X14	X15	Class
417	1	33.5	1.75	.	.	253	857	?
418	1	27.67	13.75	.	.	487	500	?
419	2	58.42	21	.	.	0	6700	?
.	?
.	?
688	2	25.25	13.5	.	.	200	1	?
689	1	17.92	0.205	.	.	280	750	?
690	1	35	3.375	.	.	0	0	?

2.1. Langkah-Langkah Klasifikasi

Penelitian ini menggunakan *tools* WEKA 3.8.3 dalam melakukan klasifikasi. WEKA adalah sebuah perangkat lunak yang memiliki banyak algoritma *machine learning* untuk keperluan data mining. Weka juga memiliki banyak *tools open source* untuk pengolahan data, mulai dari *pre-processing*, *classification*, *regression*, *clustering*, *association rules*, dan *visualization* [10].



Gambar 1. Proses klasifikasi

Gambar 1 menunjukkan tahapan sistem dalam proses klasifikasi. Tahapan klasifikasi dimulai dari *Pre Processing*, klasifikasi dan evaluasi. Berikut penjelasan untuk setiap tahap:

a) *Pre Processing*

1. *Missing Value*

Dalam dataset yang digunakan terdapat *missing value*. *Missing value* merupakan suatu kondisi dimana data memiliki *value* atau nilai yang hilang sehingga menurunkan performa klasifikasi. Oleh karena itu langkah pertama dalam penelitian ini adalah melakukan imputasi terhadap *missing value* dengan menggunakan **ReplaceMissingValue** pada *tools* WEKA, yaitu dengan mengganti data yang hilang dengan nilai modus untuk atribut bertipe nominal dan nilai mean untuk atribut bertipe kontinu.

2. Ubah Format Data

Setelah melakukan imputasi pada *missing value*, tahap selanjutnya yaitu merubah format data dari format asli yaitu *excel / .xls* menjadi format *.arff* yang agar dapat diproses oleh *tools* WEKA.

```

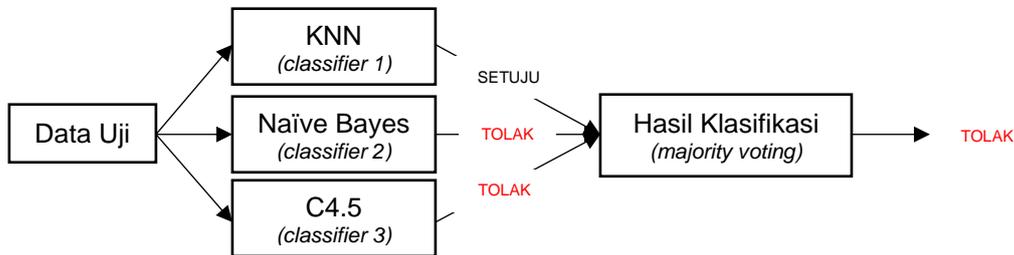
@relation          data_kredit-weka.filters.unsupervised.attribute.Remove-R1-
weka.filters.unsupervised.attribute.ReplaceMissingValues

@attribute X1 {1,2}
@attribute X2 numeric
@attribute X3 numeric
@attribute X4 {1,2,3,4}
@attribute X5 {1,2,3}
@attribute X6 {1,2,3,4,5,6,7,8,9,10,11,12,13,14}
@attribute X7 {1,2,3,4,5,6,7,8,9}
@attribute X8 numeric
@attribute X9 {1,2}
@attribute X10 {1,2}
@attribute X11 numeric
@attribute X12 {1,2}
@attribute X13 {1,2,3}
@attribute X14 numeric
@attribute X15 numeric
@attribute class {SETUJU,TOLAK}

@data
1,30.83,0,1,1,10,1,1.25,1,1,1,2,1,202,0,SETUJU
2,58.67,4.46,1,1,9,2,3.04,1,1,6,2,1,43,560,SETUJU
2,24.5,0.5,1,1,9,2,1.5,1,2,0,2,1,280,824,SETUJU
.
.
.
2,32.016495,11.25,1,1,14,8,0,2,2,0,2,1,183.356968,5200,TOLAK
1,16.5,0.125,1,1,1,1,0.165,2,2,0,2,1,132,0,TOLAK
2,27.33,1.665,1,1,14,8,0,2,2,0,2,1,340,1,TOLAK
  
```

b) Klasifikasi

Dalam melakukan klasifikasi digunakan *ensemble classifier* dengan menggunakan *majority voting* dalam menentukan kelas, yang terdiri dari algoritma KNN, *Naïve Bayes*, dan C4.5.



Gambar 2. Proses ensemble classifier

1. *K-Nearest Neighbor*

KNN (*K-Nearest Neighbor*) merupakan salah satu algoritma klasifikasi yang menggunakan kedekatan ketetanggaan data latih untuk menentukan kelas data uji [11]. Tahapan proses algoritma KNN diawali dengan menentukan parameter K kemudian menghitung nilai kedekatan ketetanggaan antara data uji dengan data latih dengan menggunakan persamaan jarak *Euclidean Distance*. Perhitungan jarak dengan *Euclidean Distance* dapat diformulasikan pada Persamaan 1.

$$d_{Eucl}(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \quad (1)$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Setelah didapatkan hasil kedekatan ketetanggaan, maka selanjutnya hasil tersebut diurutkan dari yang memiliki nilai kedekatan terkecil ke yang terbesar. Lalu diambil data yang memiliki kedekatan terkecil sebanyak K, data uji akan diklasifikasikan kelasnya sesuai dengan kelas tetangganya. Misal K=3, dengan nilai kelas tetangga diurut dari yang terdekat yaitu, SETUJU, SETUJU, TOLAK maka data uji tersebut akan dilabeli SETUJU karena tiga tetangga terdekatnya dua berlabel SETUJU dan hanya satu berlabel TOLAK, SETUJU lebih dominan.

2. *Naïve Bayes Classifier*

Naïve Bayes Classifier (NBC) adalah algoritma pengklasifikasian probabilitas sederhana yang mengaplikasikan *Teorema Bayes* dengan asumsi tidak ada ketergantungan (*independent*) yang tinggi. Salah satu kelebihan *Naïve Bayes* yaitu dalam menentukan parameter yang diperlukan dalam proses klasifikasi hanya membutuhkan data pelatihan yang kecil. Karena diasumsikan sebagai variabel independen maka hanya varians dari variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks *kovarians* [12]. Berikut ini merupakan persamaan umum *Naïve Bayes*:

$$p(C|F_1, \dots, F_n) = \frac{p(C) p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \quad (2)$$

Dimana C merepresentasikan kelas, sementara F1...Fn merepresentasikan berbagai karakteristik petunjuk yang dibutuhkan dalam melakukan klasifikasi. Maka persamaan tersebut menjelaskan bahwa peluang masuknya sebuah data dengan karakteristik tertentu ke dalam kelas C adalah peluang munculnya kelas C sebelum masuknya data tersebut dikalikan dengan peluang munculnya berbagai karakteristik data pada kelas C dibagi dengan peluang kemunculan karakteristik data secara global (*evidence*).

3. C4.5

Algoritma C4.5 yang tergolong dalam *decision tree* merupakan pengembangan algoritma ID3 yang dikembangkan oleh J. R. *Quinlan* pada tahun 1987 [12]. Dalam membangun pohon keputusan, yang pertama dilakukan adalah memilih atribut sebagai *root* atau akar, kemudian dibuat cabang untuk tiap-tiap nilai akarnya. Langkah selanjutnya adalah membagi kasus dalam cabang, kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Untuk memilih atribut akar yaitu berdasarkan nilai *gain* tertinggi dari atribut-atribut yang ada. *Gain* (S,A) adalah perolehan informasi dari atribut A relatif terhadap *output* S. Perolehan informasi didapat dari *output* data atau *variabel dependent* S, yang dikelompokkan berdasarkan atribut A, dinotasikan dengan *gain* (S, A). Berikut adalah persamaan dalam menghitung nilai *gain* [13].

$$Gain(S, A) = Entropu(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropu(S_i) \quad (3)$$

Sedangkan nilai Entropy dapat dihitung menggunakan persamaan berikut:

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

c) Evaluasi

1. Akurasi

Umumnya evaluasi kinerja dari suatu algoritma klasifikasi dapat dilakukan dengan menggunakan *confusion matrix*, yang akan menghasilkan nilai akurasi. Akurasi dalam klasifikasi merupakan persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi [12].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

2. *Holdout Method*

Dalam *holdout method*, data awal yang memiliki label kelas akan dipartisi menjadi 2 data yang dinamakan *training set* dan *test set*. Model klasifikasi kemudian dibangun menggunakan data yang ada pada *training set*, kemudian model tersebut akan diujikan ke *test set*, lalu hasil prediksi akan dibandingkan dengan kelas aktual untuk mendapatkan nilai akurasi. Proporsi data yang digunakan untuk *training set* dan *test set* tergantung pada kebutuhan analisis misalnya 50%-50%, 75%-25%, 90%-10% dimana umumnya jumlah data *training* sama atau lebih banyak dari data *testing*.

Metode ini memiliki beberapa keterbatasan. Pertama, data yang memiliki kelas berlabel yang digunakan untuk *training* akan lebih sedikit karena sebagian digunakan untuk *test set*, hasilnya tidak akan sebaik jika semua data yang memiliki kelas berlabel digunakan untuk *training*. Kedua, model yang dibangun sangat dipengaruhi oleh persentase data *training* dan *test set*. Semakin kecil ukuran *training set* maka akan semakin besar *varians* dari model, sebaliknya makin besar *training set* maka akurasi yang diestimasi dari *test set* yang berjumlah sangat kecil akan dianggap kurang *reliable*.

3. Hasil dan diskusi

Performansi algoritma *ensemble classifier* akan diuji pengaruh tingkat *imbalanced class* terhadap akurasi dan akan dibandingkan dengan akurasi dari *single classifier*.

3.1. Pengujian pengaruh tingkat Imbalanced Class terhadap akurasi

Pada pengujian ini digunakan 300 data *training* dari total 416 data *training* yang tersedia, kemudian dipecah menjadi dua bagian *training* dan *test set* sesuai dengan *holdout method*. Variabel kontrol pada pengujian ini yaitu jumlah *training set* sebanyak 250 data dan *test set* sebanyak 60 data dimana kelas untuk *test set* dibuat *balance* atau seimbang, sedangkan proporsi kelas untuk *training set* menjadi variabel bebas untuk melihat pengaruh *imbalanced class* terhadap akurasi *classifier*. Skenario pengujian pengaruh tingkat *imbalanced class* terhadap akurasi *classifier* ditunjukkan pada tabel 3.

Tabel 3. Skenario pengujian pengaruh tingkat imbalanced class terhadap akurasi classifier

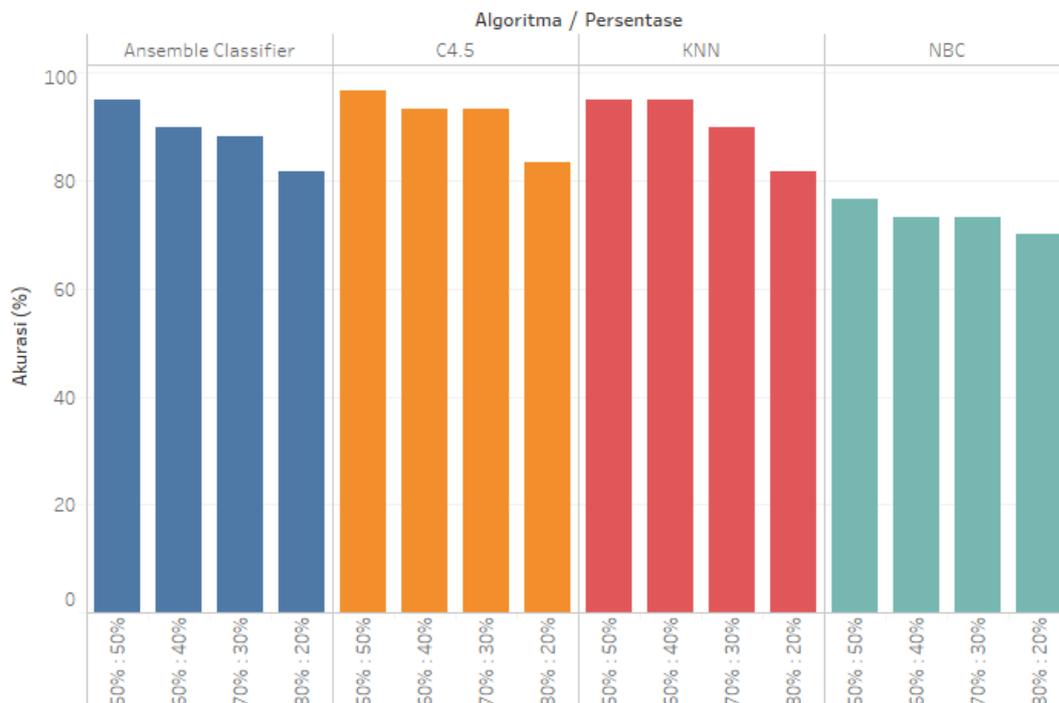
Jumlah Data Training	Kelas		Persentase Kelas TOLAK : SETUJU
	TOLAK	SETUJU	
250	125	125	50% : 50%
250	150	100	60% : 40%
250	175	75	70% : 30%
250	200	50	80% : 20%

Setelah melakukan pengujian didapatkan hasil akurasi seperti yang ditunjukkan pada Tabel 4 di bawah. Hasil klasifikasi menunjukkan adanya pengaruh tingkat *imbalanced class* terhadap nilai akurasi terhadap semua algoritma klasifikasi baik pada *single classifier* yaitu KNN, NBC, dan C4.5 maupun *ensemble classifier*.

Tabel 4. Hasil pengujian pengaruh tingkat imbalanced class terhadap akurasi classifier

Persentase Kelas TOLAK : SETUJU	Akurasi			
	KNN (K=3)	NBC	C4.5	Ensemble Classifier
50% : 50%	95%	76.67%	96.67%	95%
60% : 40%	95%	73.33%	93.33%	90%
70% : 30%	90%	73.33%	86.67%	88.33%
80% : 20%	81.67%	70%	83.33%	81.67%

Selanjutnya data pada tabel 4 digambarkan dalam bentuk diagram batang untuk melihat tingkat akurasi setiap algoritma.



Gambar 3. Hasil pengujian pengaruh tingkat imbalanced class terhadap akurasi classifier

Semakin tinggi tingkat *imbalanced class* semakin menurun pula nilai akurasinya. Untuk kasus kelas yang *balance* akurasi tertinggi didapatkan oleh algoritma C4.5 yaitu sebesar 96.67%, begitu pula untuk kasus *imbalanced class* 80% (TOLAK) : 20% (SETUJU) akurasi tertinggi

didapatkan oleh algoritma C4.5 yaitu sebesar 83.33%. Secara garis besar NBC dianggap merupakan algoritma dengan akurasi terendah dalam kasus penentuan persetujuan pemberian kredit, sedangkan C4.5 dianggap yang tertinggi. Namun dapat dilihat baik C4.5 maupun KNN mengalami penurunan akurasi yang relatif cukup besar ketika kasus *imbalanced class* 80% (TOLAK) : 20% (SETUJU), sedangkan *Ensemble Classifier* dan NBC dianggap relatif lebih stabil.

Ensemble Classifier memperoleh akurasi terbaik pada kasus kelas yang *balance* yaitu sebesar 95%, dan akurasi terendah diperoleh untuk kasus *imbalanced class* 80% (TOLAK) : 20% (SETUJU) yaitu sebesar 81.67%. *Ensemble Classifier* dianggap relatif cukup baik dalam menangani kasus persetujuan pengajuan kredit dengan rata-rata akurasi sebesar 88.75%.

3.2. Implementasi Hasil

Karena hasil pengujian menunjukkan algoritma *Ensemble Classifier* relatif cukup baik dalam menangani kasus *imbalanced class*, maka algoritma tersebut selanjutnya digunakan untuk melakukan klasifikasi pada data *testing* sebanyak 274 data dengan menggunakan keseluruhan data *training* yaitu 416 data untuk membangun model klasifikasi, yang terdiri dari 231 data dengan kelas = TOLAK dan 185 data dengan kelas = SETUJU.

Hasil klasifikasi *Ensemble Classifier* terhadap data 274 data *testing* yaitu sebanyak 109 nasabah dinyatakan SETUJU dan sebanyak 165 nasabah dinyatakan pengajuan kreditnya di TOLAK.

Tabel 5. Contoh Hasil Klasifikasi Algoritma Ensemble Classifier Terhadap Data Testing

Nasabah	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	aktual_class	C4.5	NBC	KNN	Ensemble Classifier
417	1	33.5	1.75	1	1	11	2	4.5	1	1	4	1	1	253	857	?	SETUJU	SETUJU	SETUJU	SETUJU
418	1	27.67	13.75	1	1	10	1	5.75	1	2	0	1	1	487	500	?	SETUJU	TOLAK	SETUJU	SETUJU
419	2	58.42	21	1	1	4	3	10	1	1	13	2	1	0	6700	?	SETUJU	SETUJU	SETUJU	SETUJU
420	2	20.67	1.835	1	1	9	1	2.085	1	1	5	2	1	220	2503	?	SETUJU	SETUJU	SETUJU	SETUJU
421	1	26.17	0.25	1	1	4	3	0	1	2	0	1	1	0	0	?	SETUJU	TOLAK	TOLAK	TOLAK
422	1	21.33	7.5	1	1	13	1	1.415	1	1	1	2	1	80	9800	?	SETUJU	SETUJU	SETUJU	SETUJU
423	1	42.83	4.625	1	1	9	1	4.58	1	2	0	2	3	0	0	?	SETUJU	TOLAK	TOLAK	TOLAK
424	1	38.17	10.125	1	1	11	1	2.5	1	1	6	2	1	520	196	?	SETUJU	SETUJU	SETUJU	SETUJU
425	1	20.5	10	2	2	1	1	2.5	1	2	0	2	3	40	0	?	TOLAK	TOLAK	TOLAK	TOLAK

4. Kesimpulan

Ketidakseimbangan kelas pada dataset merupakan suatu hal yang sering ditemui pada domain dunia nyata (*real world problem*) terutama dalam kasus klasifikasi data mining. Berdasarkan penelitian yang telah dilakukan *ensemble classifier* dianggap relatif cukup baik dalam menangani kasus persetujuan pengajuan kredit dengan rata-rata akurasi sebesar 88.75%. Sehingga *ensemble classifier* dengan *majority voting* dapat menjadi alternatif solusi dalam menangani kasus *imbalanced class*. Untuk penelitian selanjutnya diparapkan penelitian dapat dilakukan dengan menggunakan metode *ensemble* lainnya dan diujikan pada dataset yang lebih beragam, misalnya implementasi *ensemble classifier* pada kasus *big data* yang berpotensi mengalami *imbalanced class*.

5. Notasi

$dEucl(p, q)$

p : nilai dari data latih

q : nilai dari data uji

i : data ke- i

n : banyaknya data

$Gain(S, A)$

S : himpunan kasus

A : Atribut

n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke- i

$|S|$: jumlah kasus dalam S

$Entropy(S)$

S : himpunan kasus

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

$Accuracy$

$True Positive (TP)$: jumlah record positif yang diklasifikasikan sebagai positif

False Positive (FP) : jumlah record negatif yang diklasifikasikan sebagai positif
True Negative (TN) : jumlah record negatif yang diklasifikasikan sebagai negatif
False Negative (FN) : jumlah record negatif yang diklasifikasikan sebagai positif

Referensi

- [1] L. Marpaung, *Pemberantasan dan pencegahan tindak pidana terhadap perbankan*. Djambatan, 2003.
- [2] R. Lihani, N. Ngadiman, and N. Hamidi, "Analisis Manajemen Kredit Guna Meminimalkan Risiko Kredit (Studi pada PD BPR BKK Tasikmadu Karanganyar)," *J. Pendidik. Ekon. Univ. Sebel. Maret*, vol. 1, no. 3, p. 13534, 2013.
- [3] E. Turban, J. E. Aronson, and T. P. Liang, "Decision Support Systems and Intelligent System,(Sistem Pendukung Keputusan dan Sistem Cerdas) Ed. 7. Jld. 2," 2005.
- [4] E. N. Jannah, R. W. Herlambang, and M. R. Pangow, "Penentuan Persetujuan Pemberian Kredit Bank Menggunakan Metode Hybrid SVM - Penelusuran Google," *Data Mining Contest - Gemastik*, 2011. .
- [5] H. Marcos and I. Hidayah, "Implementasi Data Mining Untuk Klasifikasi Nasabah Kredit Bank 'X'Menggunakan Classification Rule," *Semin. Nas. Teknol. Inf. Dan Multimed*, pp. 1–7, 2014.
- [6] Y. Pristyanto, "Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset," *J. Teknoinfo*, vol. 13, no. 1, pp. 11–16, 2019.
- [7] N. V Chawla, N. Japkowicz, and A. Kotcz, "Special issue on learning from imbalanced data sets," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 1–6, 2004.
- [8] P. N. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining, Boston: Person Education." Inc, 2006.
- [9] N. F. F. Da Silva, E. R. Hruschka, and E. R. Hruschka Jr, "Tweet sentiment analysis with classifier ensembles," *Decis. Support Syst.*, vol. 66, pp. 170–179, 2014.
- [10] "Documentation - Weka Wiki." [Online]. Available: <https://waikato.github.io/weka-wiki/documentation/>. [Accessed: 01-Sep-2020].
- [11] D. A. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Appl. Comput. Informatics*, vol. 12, no. 1, pp. 90–108, 2016.
- [12] J. Han, M. Kamber, and J. Pei, "Data Mining: Concept and Technique," *Morgan Kaufmann*, 2016. .
- [13] E. T. L. Kusriani, "Algoritma data mining," *Yogyakarta Andi Offset*, 2009.