

PERBANDINGAN METODE RANDOM FOREST DAN NAÏVE BAYES UNTUK KLASIFIKASI DEBITUR BERDASARKAN KUALITAS KREDIT

Bonggo Bawono¹⁾, Rochdi Wasono²⁾.

¹FMIPA, Universitas Muhammadiyah Semarang
email: bonggosaptaji@gmail.com

Abstract

Pertumbuhan kredit dari tahun 2016 – 2018 mengalami peningkatan. Peningkatan tersebut bersumber dari semua jenis penggunaan kredit, baik modal kerja, investasi, maupun konsumsi. Namun, tidak semua kredit dapat dikembalikan secara sempurna dan tepat waktu artinya akan muncul suatu resiko yang dikenal dengan resiko kredit dimana resiko kredit dapat terjadi pada setiap bank. Masalah resiko kredit dapat diatasi dengan klasifikasi. Metode klasifikasi yang pernah digunakan adalah Random Forest dan Naive Bayes, namun Random Forest dan Naive Bayes memiliki kelebihan dan kekurangan masing-masing oleh karena itu dilakukan perbandingan metode dengan membandingkan nilai Akurasi. Hasil yang didapat Random Forest memiliki hasil akurasi tertinggi yakni sebesar 98,16% disusul Naive Bayes 95,93%. Sehingga dengan Random Forest nasabah yang memiliki kredit bermasalah hanya 3,22%

Keywords: *Debitur, klasifikasi, Random Forest, Naive Bayes, Kualitas Kredit*

1. PENDAHULUAN

Survei Perbankan Bank Indonesia pada Desember 2018 memperoleh hasil permintaan kredit baru triwulan IV-2018 meningkat 21,2% pada triwulan sebelumnya yakni 71,7%. Pertumbuhan kredit dari tahun 2016 – 2018 mengalami peningkatan. Namun, pada faktanya, tidak semua kredit dapat dikembalikan secara sempurna dan tepat waktu artinya akan muncul suatu resiko yang dikenal dengan resiko kredit dimana resiko kredit dapat terjadi pada setiap bank. Resiko kredit dapat terjadi akibat ketidakmampuan nasabah dalam membayar kewajibannya dalam jangka waktu yang telah ditentukan dalam perjanjian pemberian kredit oleh pihak Bank kepada nasabah yang akan mengakibatkan kredit bermasalah (*non performing loan*) yang semakin besar (Hamonangan dkk, 2009).

Berdasarkan Peraturan Bank Indonesia No. 8/19/ 2006 Indonesia tentang Kualitas Aktiva Produktif dan Pembentukan Penyisihan Penghapusan Aktiva Produktif Bank Perkreditan Rakyat , membagi kualitas kredit dua golongan yakni kredit tidak bermasalah dan kredit bermasalah. Kredit bermasalah adalah kredit dengan kualitas kurang lancar, diragukan dan macet (Riyadi, 2006). Menurut Siswanto Sutojo (2002) Kredit *non performing* pada umumnya merupakan kredit yang pembayaran angsuran pokok dan atau bunganya telah lewat sembilan puluh hari lebih setelah jatuh tempo, atau kredit yang pembayarannya secara tepat waktu sangat diragukan. Kredit bermasalah perolehan laba dan pengaruh buruk bagi profitabilitas bank dan dapat menyebabkan kebangkrutan.

Dalam menghadapi masalah resiko kredit yang dialami oleh Industri perbankan saat ini salah satunya dapat diatasi dengan mengidentifikasi dan memprediksi nasabah dengan baik sebelum memberikan pinjaman dengan cara memperhatikan data historis pinjaman. Oleh karena itu klasifikasi resiko kredit dalam perbankan memiliki peran yang penting. Apabila pengklasifikasian debitur mengalami kesalahan, maka salah satu dampak yang ditimbulkan adalah kredit bermasalah. Dalam menyelesaikan permasalahan

klasifikasi, penggunaan metode atau teknik bertujuan untuk mempermudah proses klasifikasi. Beberapa teknik yang digunakan dalam kasus klasifikasi yaitu *decision tree*, *classification and association rule*.

Penelitian terdahulu tentang klasifikasi kredit bank telah dilakukan oleh Akhmad Syukron dan Agus Subekti (2018) menggunakan teknik *data mining* dengan menggunakan Random Forest dan Random Over-Under Sampling. Hasil menunjukkan, metode *Random Forest* memiliki nilai akurasi yang lebih baik dibandingkan dengan beberapa metode lainnya. Sedangkan klasifikasi dengan penerapan metode *Random Over-under sampling Random Forest* dapat meningkatkan kinerja akurasi sebesar 14,1%. Selain itu, Oktanisa dan Supianto (2017) melakukan penelitian untuk klasifikasi Bank Direct Marketing dengan metode *Support Vector Machine*, *AdaBoost*, *Naïve Bayes*, *Constant*, *KNN*, *Tree*, *Random Forest*, *Stochastic Gradient Descent*, dan *CN2 Rule* dan akurasi terbaik diperoleh oleh model *Tree*, *Constant*, dan *Naive Bayes*.

Random Forest adalah pengembangan dari metode CART, yaitu dengan menerapkan metode *bootstrap aggregating (bagging)* dan random feature selection (Breiman 2001). Metode ini merupakan metode pohon gabungan. Dalam random forest, banyak pohon ditumbuhkan sehingga terbentuk suatu hutan (forest), kemudian analisis dilakukan pada kumpulan pohon tersebut. Penggunaan *Random Forest* dapat menghindari *overfitting* pada sebuah set data saat mencapai akurasi yang maksimum. Namun terdapat *Class imbalance* atau ketidak-seimbangan kelas yang merupakan salah satu permasalahan pada data mining. Hal ini terjadi pada saat kelas minoritas jauh lebih kecil atau lebih jarang dari kelas mayoritas (Syukron dan Subekti, 2018).

Naïve Bayes adalah metode yang digunakan dalam statistika untuk menghitung peluang dari suatu hipotesis, *Naïve Bayes* menghitung peluang suatu label berdasarkan pada atribut yang dimiliki dan menentukan label yang memiliki peluang paling tinggi (Lu, Ling, & Huang, 2003). Metode *Naïve Bayes* juga memiliki beberapa keunggulan seperti mudah serta biaya perhitungan kecil (Wu & Kumar, 2009), dapat menangani data missing (Kang, Yoo, & Han, 2012). Namun metode *Naïve Bayes* memiliki asumsi independensi atribut (Zhang, Liu, & Almpandis, 2017).

Penelitian yang sudah dilakukan masing-masing metode memiliki kekurangan dan kelebihan masing-masing oleh karena itu penulis akan melakukan perbandingan terhadap metode Random Forest dan *Naïve bayes* untuk klasifikasi debitur berdasarkan kualitas kredit dengan membandingkan nilai akurasi.

2. METODE PENELITIAN DATA

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari salah satu Bank. Data yang memiliki 10 atribut yang terdiri dari Tanggal Akad, Tanggal Jatuh Tempo, Plafond, Saldo September 2017, Tunggakan Pokok, Tunggakan Bunga, Suku Bunga, Jaminan, Nilai Jaminan, Jangka Waktu (bulan) dan satu label yakni Tingkat Kolektibilitas. Jumlah keseluruhan dataset yang digunakan ada 6320 data.

No	Variabel	Jenis Data	Keterangan
1	Tanggal Akad	Date	Atribut
2	Tanggal Jatuh Tempo	Date	Atribut
3	Plafond	Rasio	Atribut
4	Saldo September 2017	Rasio	Atribut
5	Tunggakan Pokok	Rasio	Atribut

6	Tunggakan Bunga	Rasio	Atribut
7	Suku Bunga	Rasio	Atribut
8	Jangka Waktu (bulan)	Rasio	Atribut
9	Nilai Jaminan	Rasio	Atribut
10	Jaminan	Ordinal	Atribut
11	Tingkat Kolektibilitas	Ordinal	Label

Langkah Penelitian

Penelitian dilakukan dengan melakukan *preprocessing* data yakni untuk melakukan pemilihan atribut yang akan digunakan. Selanjutnya dilihat terdapat data *missing* atau tidak. Lalu data dibagi menjadi dua yakni data training dan data testing. Data training adalah data untuk membuat model dan data testing untuk menguji model. Pembagian data dengan menggunakan X-Cross Fold Validation. Pada data training terdapat dua metode yang digunakan yakni Random Forest dan Naïve Bayes berikut langkah kedua metode:

Random Forest

Pada gugus data yang terdiri atas n amatan dan p peubah penjelas, prosedur untuk melakukan *random forest* sebagai berikut:

1. Lakukan penarikan contoh acak berukuran n dengan pemulihan pada gugus data. Tahap ini adalah tahapan *bootstrap*.
2. Dengan menggunakan contoh *bootstrap*, pohon dibangun sampai mencapai ukuran maksimum (tanpa pemangkasan). Pembangunan pohon dilakukan dengan menerapkan *random feature selection* pada setiap proses pemilihan pemilah, yaitu m peubah penjelas dipilih secara acak dimana $m \ll p$, lalu pemilah terbaik dipilih berdasarkan m peubah penjelas tersebut.
3. Ulangi langkah 1 dan 2 sebanyak k kali, sehingga terbentuk sebuah hutan yang terdiri atas k pohon.

Naïve Bayes

Tahapan *Naïve Bayes* dapat dilakukan sebagai berikut:

1. Menghitung peluang masing-masing label $P(C_i)$
2. Menentukan peluang X atau kategori dari atribut
3. Menghitung peluang kategori dari atribut yang sama dengan label yang sama $P(X|C_i)$
4. Kalikan semua nilai hasil sesuai dengan data yang dicari labelnya
5. Tentukan nilai label pada atribut tertentu

Setelah mendapatkan model lalu dilakukan pengujian model dengan menggunakan data testing dan dilakukan proses perhitungan nilai akurasi. Evaluasi pengujian yang dilakukan terhadap metode klasifikasi Random Forest dan *Naïve Bayes*. Penelitian ini menggunakan tingkat akurasi yang dihasilkan metode tersebut untuk klasifikasi debitur berdasarkan kualitas kredit Perhitungan akurasi dapat dilihat ditabel 1.

Tabel 1. Perhitungan Akurasi

Label	Kenyataan Benar	Kenyataan Salah
Prediksi Benar	TP (True Positive)	FP (False Positive)
Prediksi Salah	FN (False Negative)	TN (True Negative)

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \times 100$$

3. HASIL PENELITIAN

Hasil

Pada pengujian ini semua *dataset* sejumlah 6320 data akan dijadikan data pelatihan dan data pengujian. Pengujian dilakukan dengan menggunakan aplikasi *Rapidminer*. Hasil dari model pengujian ini dapat dilihat pada tabel berikut:

No	Metode	Prediksi Benar	Prediksi Salah	Akurasi
1	Random Forest	6204	116	98,16%
2	Naive Bayes	6063	257	95,93%

Secara terperinci dapat dijelaskan berdasarkan prediksi dengan menggunakan naive bayes terdapat 5856 nasabah dikategorikan kredit tidak bermasalah serta 464 nasabah mengalami kredit bermasalah sedangkan prediksi dengan menggunakan random forest terdapat 6117 nasabah dikategorikan kredit tidak bermasalah serta 203 nasabah mengalami kredit bermasalah.



Selain itu Naive Bayes mampu memprediksi kredit bermasalah dengan tepat sebanyak 261 kasus, memprediksi salah pada kasus kredit tidak bermasalah sebanyak 203 kasus, memprediksi salah pada kasus kredit bermasalah sebanyak 54 kasus dan memprediksi kredit tidak bermasalah secara tepat sebanyak 5802 kasus. Sementara itu, Random Forest mampu memprediksi kredit bermasalah dengan tepat sebanyak 201 kasus, memprediksi salah pada kasus kredit tidak bermasalah sebanyak 2 kasus, memprediksi salah pada kasus kredit bermasalah sebanyak 114 kasus dan memprediksi kredit tidak bermasalah secara tepat sebanyak 6003 kasus.

accuracy: 95.93% +/- 0.65% (mikro: 95.93%)			
	true Kredit Bermasalah	true Kredit Tidak Bermasalah	class precision
pred. Kredit Bermasalah	261	203	56.25%
pred. Kredit Tidak Bermasalah	54	5802	99.08%
class recall	82.86%	96.62%	

Naive Bayes

accuracy: 98.16% +/- 0.70% (mikro: 98.16%)

	true Kredit Bermasalah	true Kredit Tidak Bermasalah	class precision
pred. Kredit Bermasalah	201	2	99.01%
pred. Kredit Tidak Bermasalah	114	6003	98.14%
class recall	63.81%	99.97%	

Random Forest

Berdasarkan Kurva ROC dihasilkan dari tarik ulur antara sensitivitas dan spesifisitas pada berbagai titik potong. Dari prosedur ROC ini kita akan mendapatkan nilai *Area Under Curve* (AUC) yakni 0,98 yang berdasarkan kategori AUC yakni sangat baik sedangkan Berdasarkan Kurva ROC Random Forest dihasilkan dari tarik ulur antara sensitivitas dan spesifisitas pada berbagai titik potong. Dari prosedur ROC ini kita akan mendapatkan nilai *Area Under Curve* (AUC) yakni 0,974 yang berdasarkan kategori AUC yakni sangat baik.



4. KESIMPULAN

Kesimpulan yang diperoleh pada penelitian ini adalah dari hasil perbandingan klasifikasi dengan Random Forest dan Naive bayes pada studi kasus klasifikasi debitur berdasarkan kualitas kredit menyatakan bahwa klasifikasi Random Forest merupakan metode klasifikasi yang memiliki tingkat akurasi paling tinggi untuk klasifikasi kualitas kredit yaitu mencapai 98,16% disusul Naive Bayes 95,93%. Sehingga dengan Random Forest nasabah yang memiliki kredit bermasalah hanya 3,22%.

5.REFERENSI

Hamonangan, Reynaldo dan Hasan Sakti Siregar. 2009. *Pengaruh Capital Adequancy Ratio, Debt to Equity Ratio, Non Performing Loan, Operating Ratio dan Loan to Deposit Ratio terhadap Return On Equity (ROE) Perusahaan Perbankan yang Terdaftar di Bursa Efek Indonesia*, Jurnal Akuntansi 13, Universitas Sumatera Utara, Medan.

Indonesia, G. B. 2006. *Peraturan Bank Indonesia Nomor: 8/19/PBI/2006 tentang Kualitas Aktiva Produktif dan Pembentukan Penyisihan Penghapusan Aktiva Produktif Bank Perkreditan Rakyat*. Jakarta.

Kang, H., Yoo, S. J., & Han, D. 2012. Expert Systems with Applications Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems With Applications*, 39(5), 6000–6010.

- <https://doi.org/10.1016/j.eswa.2011.11.107>
- Lu, J., Ling, C. X., & Huang, J. (2003). Comparing Naive Bayes , Decision Trees , and SVM with AUC and Accuracy. *The Third IEEE International Conference on DataMining*, 11–14.
- Riyadi, Selamet. 2006. *Banking Assets and Liability Management* (Edisi Ketiga). Jakarta: lembaga Penerbit FEUI.
- Sutojo, Siswanto. 2002. *Menangani Kredit Bermasalah: konsep, teknik, dan kasus*. Jakarta: Gramedia.
- Syukron dan Subekti, 2018. *Penerapan Metode Random Over-Under Sampling dan Random Forest untuk Klasifikasi Penilaian Kredit*. *Jurnal Informatika* Vol.5 No.2
- Oktanisa dan Supianto, 2017. *Perbandingan Teknik Klasifikasi dalam Data Mining untuk Bank Direct Marketing*. *Jurnal Teknologi Informasi dan Ilmu Komputer* Vol. 5, No. 5, Oktober 2018, hlm. 567-576. <https://doi.org/10.25126/jtiik20185958>
- Wu, X., & Kumar, V. 2009. *The Top Ten Algorithm in Data Mining*. Taylor & Francis Group, LLC.
- Zhang, C., Liu, C., Zhang, X., & Alpanidis, G. 2017. An Up-to-Date Comparison of State-of-the-Art Classification Algorithms. *Expert Systems With Applications*. <https://doi.org/10.1016/j.eswa.2017.04.003>

