

Peningkatan Akurasi Pada Modified K-NN Untuk Klasifikasi Pengajuan Kredit Koperasi Dengan Menggunakan Algoritma Genetika

Bambang Ismanto, Nurul Amalia
STMIK Widya Pratama Pekalongan
Jl. Patriot No.25 Pekalongan Jawa Tengah
E-mail: bams.stmikwp@gmail.com, amalia.0121@gmail.com

RINGKASAN

Pemberian kredit kepada konsumen atau calon nasabah melewati beberapa proses. Proses ini dimulai dari pengajuan kredit kemudian dari proses pengajuan kredit tersebut akan dilakukan proses analisis pemberian kredit. Kesalahan mengidentifikasi kredit menyebabkan risiko kredit yang mengarah pada hilangnya pendapatan dan memperluas kredit untuk risiko kredit yang bertipe buruk yaitu ancaman bagi profitabilitas. Teknik klasifikasi data mining dapat digunakan untuk menentukan risiko kredit. Algoritma klasifikasi banyak digunakan oleh peneliti dengan dataset dan hasil yang berbeda pula. k-NN merupakan yang terbaik untuk data dengan 10 atribut dibandingkan hasil Naive Bayes, SVM dan C-45. Penelitian dalam hal perbaikan kNN sudah banyak dilakukan oleh banyak peneliti, baik dari segi akurasi maupun dari segi optimasi nilai k. Modified kNN (MKNN) merupakan salah satu algoritma hasil penelitian untuk meningkatkan nilai akurasi. Perbaikan yang dilakukan yaitu dengan menambahkan perhitungan nilai validity dalam perhitungan nilai bobot pada kNN tradisional yang digunakan untuk mengatasi masalah outlier. Namun algoritma MKNN masih memiliki kelemahan utama, yaitu penentuan nilai k, biaya komputasi cukup tinggi, rentan terhadap variabel yang non-informatif. Salah satu algoritma yang digunakan untuk mengatasi masalah optimasi yaitu Algoritma Genetika. Algoritma Genetika dapat menentukan nilai k secara otomatis dan dapat meningkatkan nilai akurasi serta mengurangi kompleksitas. Dari hasil eksperimen yang dilakukan dengan menerapkan Algoritma Genetika pada Algoritma Modified K-Nearest Neighbour diperoleh hasil akurasi sebesar 96,47%, recall sebesar 96,47% dan precision sebesar 91,25%.

Kata Kunci : Klasifikasi pengajuan kredit, Modified kNN, Algoritma genetik

1. PENDAHULUAN

1.1 Latar Belakang

Koperasi merupakan badan usaha yang melakukan transaksi kredit dalam bentuk pinjaman anggota. KSP menjadi salah satu sumber pembiayaan utama bagi pelaku usaha mikro dan kecil mudah dijangkau dengan syarat yang tidak berbelit. KSP merupakan industri yang penuh dengan risiko, salah satunya adalah risiko kredit macet sebagai risiko kerugian yang karena nasabah tidak memenuhi kewajiban membayar tagihan pada tanggal jatuh tempo yang telah ditentukan.

Masalah kredit macet yang di Indonesia menduduki posisi tertinggi yakni 55%. Tingginya kredit macet di Indonesia oleh karena Bank-Bank di Indonesia dalam pemberian kredit kurang patuh terhadap prinsip-prinsip kehati-hatian.

Penerapan prinsip kehati-hatian dalam memberikan pinjaman akan memperkecil risiko kredit yang mungkin dialami oleh koperasi. Prinsip kehati-hatian ini dilakukan dengan

memprediksi, mengidentifikasi, mengukur, dan mengendalikan bentuk kerugian yang dapat terjadi dari risiko kredit macet dengan cara menganalisa data pinjaman sebelumnya. Oleh karena itu peran klasifikasi dalam menganalisa risiko kredit perbankan sangatlah penting. Salah satu dampak yang dapat terjadi jika pengklasifikasian risiko kredit mengalami kesalahan adalah kredit macet. Kredit macet ini dapat mengakibatkan bank tersebut mengalami kebangkrutan. Teknologi informasi dapat digunakan sebagai alat bantu dalam mengidentifikasi dan mengukur potensi risiko kredit sehingga koperasi dapat menganalisa riwayat peminjaman, serta memprediksi kelancaran kredit peminjam sebagai acuan dalam menentukan besarnya pinjaman yang akan diberikan.

Teknik klasifikasi data mining dapat digunakan untuk menentukan risiko kredit. Terdapat banyak teknik klasifikasi data mining seperti yang tercantum dalam. Klasifikasi membutuhkan data training untuk mengenali pola tertentu dari data dengan label atau hasil akhir. Kemudian pola

tersebut dipakai untuk menentukan label yang belum diketahui dari data baru.

Algoritma klasifikasi banyak digunakan oleh peneliti dengan dataset (Kusrini, 2009) dan hasil yang berbeda pula. Penelitian mengenai Pemilihan algoritma klasifikasi data mining untuk menentukan resiko kredit maupun dalam meningkatkan akurasi pernah dilakukan oleh beberapa peneliti, diantaranya:

Penelitian Siallagan (Siallagan, 2015) dalam pencarian nasabah untuk pemberian kredit digunakan algoritma C 4.5 dengan mempertimbangkan 8 atribut, dari model yang diusulkan didapatkan hasil akurasi sebesar 87,36%.

Pada penelitian Abdussomad (Abdussomad, 2014) algoritma Particle Swarm Optimization (PSO) digunakan untuk meningkatkan akurasi pada algoritma C4.5 dalam memprediksi kelayakan kredit. PSO digunakan pada tahap preprocessing, dari hasil evaluasi yang telah dilakukan didapat accuracy sebesar 91,18%.

Pada penelitian yang berbeda, para peneliti juga telah menerapkan teknik lain untuk meningkatkan akurasi pada klasifikasi pengajuan kredit, Penelitian Ester Arisawati (Arisawati, 2017) untuk meningkatkan akurasi pada algoritma k-Nearest Neighbor ditambahkan algoritma genetik yang digunakan untuk optimasi atribut. Berdasarkan hasil pengujian diperoleh akurasi yang lebih tinggi yaitu 68,56% dibandingkan akurasi dari algoritma k-Nearest Neighbor sebesar 54,32%.

Berdasarkan penelitian yang dilakukan Nandang Iriadi (Iriadi, 2012) mengenai komparasi tiga metode klasifikasi yaitu Naive Bayes, Neural Network dan K-Nearest Neighbour dalam melakukan klasifikasi peminjam pada Koperasi Serba Usaha diperoleh tingkat keakuratan K-Nearest Neighbour mendapatkan hasil 93,00% lebih baik dibandingkan dengan metode Naive Bayes sebesar 90,33% dan Neural Network sebesar 85,67%.

Metode K-Nearest Neighbor adalah metode klasifikasi yang dapat dimanfaatkan untuk mengevaluasi dan menentukan resiko kredit macet (Leidiyana, 2013). Algoritma K-Nearest Neighbor (KNN) merupakan algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran, yang diambil dari k tetangga terdekatnya (nearest neighbors). Dengan k merupakan banyaknya tetangga terdekat yang ditentukan oleh user.

Kelebihan dari kNN tangguh terhadap data latih yang *noisy* dan efekti digunakan pada data pelatihan yang besar. Sedangkan kekurangannya

yaitu penentuan nilai k, biaya komputasi cukup tinggi, rentan terhadap variabel yang non-informatif (Bathia, 2010).

Penelitian dalam hal perbaikan kNN sudah banyak dilakukan oleh banyak peneliti, baik dari segi akurasi maupun dari segi optimasi nilai k. Modified kNN (MKNN) merupakan salah satu algoritma hasil penelitian untuk meningkatkan nilai akurasi. Perbaikan yang dilakukan yaitu dengan menambahkan perhitungan nilai validity dalam perhitungan nilai bobot pada kNN tradisional yang digunakan untuk mengatasi masalah outlier. Namun algoritma MKNN masih memiliki kelemahan utama, yaitu penentuan nilai k, biaya komputasi cukup tinggi. Salah satu algoritma yang digunakan untuk mengatasi masalah optimasi yaitu Algoritma Genetika. Algoritma Genetika dapat mencari nilai k yang efektif dan dapat meningkatkan nilai akurasi serta mengurangi kompleksitas.

Berdasarkan masalah tersebut, penulis mengusulkan penerapan Algoritma Genetika pada algoritma MKNN, selanjutnya algoritma tersebut dinamakan algoritma Genetic Modified k Nearest Neighbor (GMKNN).

2. METODE PENELITIAN

2.1 Pengumpulan Data

Penelitian ini menggunakan dataset pengajuan kredit pada Koperasi Rahman Pemalang. Dataset ini memiliki 12 atribut, dengan 1 atribut label dengan jumlah data 866 record, Sebagai label atau kelas pada dataset ini adalah atribut Status Kredit. Dataset yang masih berisi data informasi pendukung perlu dilakukan pengolahan dengan membersihkan data informasi pendukung tersebut. Beberapa atribut dalam dataset ini tidak digunakan dan akan dihapus, yaitu Nama Nasabah. Dataset ini juga memiliki beberapa nilai yang tidak sesuai seperti missing value. Data tersebut muncul karena berbagai kemungkinan seperti kesalahan prosedur, dan lain-lain. Data seperti ini perlu dihilangkan karena dapat menyebabkan hasil klasifikasi yang tidak optimal. Pada penelitian ini data missing value akan dihapus.

2.2 Eksperimen

Penelitian ini akan menggunakan sebuah metode optimasi Algoritma Genetika. Algoritma Genetika digunakan untuk mengatasi permasalahan penentuan nilai k pada Algorithm Modified KNN sehingga dapat menentukan nilai k yang lebih baik dan hasil klasifikasi yang lebih baik. Berikut diagram alir desain eksperimen pada penelitian ini.

Tahap pertama yaitu menentukan parameter yang diperlukan dalam Algoritma Genetika yaitu Jumlah Populasi, Jumlah Gen, Probabilitas Crossover, Probabilitas Mutasi dan Jumlah iterasi. Parameter ini digunakan dalam proses perhitungan Algoritma Genetika.

Tahap kedua merupakan tahapan dari Algoritma Genetika. Dimulai dari melakukan inialisasi populasi awal atau pembentukan populasi awal. Populasi awal dibentuk dari dataset. Kemudian menghitung nilai fitness menggunakan nilai validitas. Hasil dari fitness kemudian akan digunakan untuk mencari kromosom induk yang selanjutnya akan dilakukan crossover atau kawin silang. Populasi hasil dari crossover akan dilakukan mutasi untuk perubahan kromosomnya. Selanjutnya akan dilakukan proses perhitungan fitness lagi untuk mencari kromosom terbaik dalam proses elitisme yang hasilnya akan digunakan dalam proses iterasi selanjutnya pada Algoritma Genetika. Tahap kedua ini akan diulang sebanyak nilai iterasi yang ditentukan atau nilai fitness terpenuhi. Jika kriteria sudah terpenuhi, maka hasil dari proses ini akan menghasilkan nilai k yang akan digunakan dalam proses klasifikasi Modified KNN.

Tahap ketiga yaitu proses klasifikasi menggunakan Modified KNN. Dimulai dari menentukan nilai k (tahap 2) dan menentukan data awal yaitu dataset pengajuan kredit. Selanjutnya akan dihitung jarak antar data dengan menggunakan euclidean distance. Kemudian langkah selanjutnya yaitu menghitung nilai k, menghitung validasi dan menghitung bobot. Proses ini akan menghasilkan model pada klasifikasi pengajuan kredit.

Tahap keempat yaitu melakukan evaluasi hasil klasifikasi yang dihasilkan dari proses klasifikasi Modified KNN. Evaluasi dalam penelitian ini menggunakan Confusion Matrix.

3. HASIL DAN PEMBAHASAN

3.1 Klasifikasi menggunakan Algoritma K-Nearest Neighbour

Percobaan klasifikasi pengajuan kredit menggunakan algoritma K-Nearest Neighbour menggunakan Aplikasi Rapid Miner. Pada penelitian ini menggunakan data 823 record. Hasil klasifikasi diuji dengan menggunakan 10-fold cross validation. Dari beberapa kali percobaan dengan menggunakan Rapid Miner, diperoleh nilai akurasi tertinggi pada nilai $k = 1$.

Dari hasil eksperimen yang dilakukan dengan menggunakan Algoritma K-Nearest Neighbour standar diperoleh hasil akurasi sebesar 89,91 %, recall sebesar 90,67 % dan precision sebesar 72,91 %.

3.2 Klasifikasi menggunakan Algoritma K-Nearest Neighbour dengan Algoritma Genetika

Percobaan klasifikasi pengajuan kredit menggunakan algoritma K-Nearest Neighbour menggunakan Aplikasi Rapid Miner. Pada penelitian ini menggunakan data 823 record. Hasil klasifikasi diuji dengan menggunakan 10-fold cross validation. Dari beberapa kali percobaan dengan menggunakan Rapid Miner, diperoleh nilai akurasi tertinggi pada nilai $k = 2$.

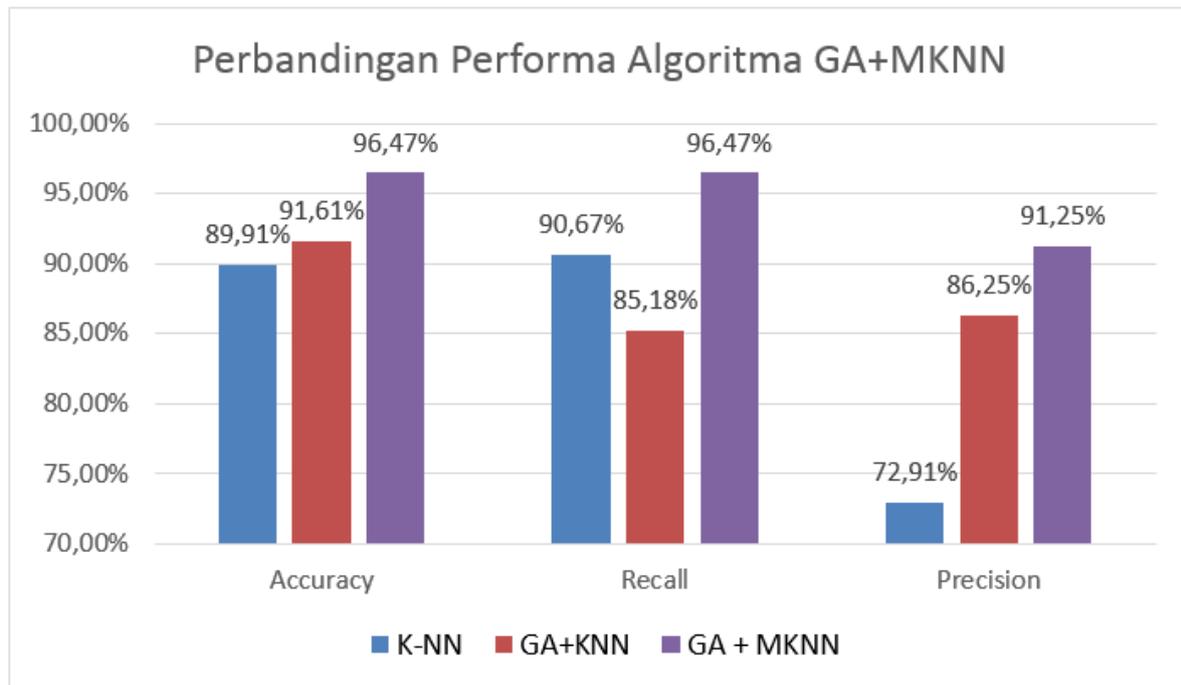
Dari hasil eksperimen yang dilakukan dengan menggunakan Algoritma K-Nearest Neighbour standar diperoleh hasil akurasi sebesar 91,61 %, recall sebesar 85,18 % dan precision sebesar 86,25 %.

3.3 Klasifikasi menggunakan Algoritma Modified K-Nearest Neighbour dengan Algoritma Genetika

Percobaan selanjutnya dilakukan klasifikasi pengajuan kredit menggunakan Algoritma Modified K-Nearest Neighbour dengan optimasi nilai K menggunakan Algoritma Genetika. Percobaan ini menggunakan sebuah program yang dibuat dengan menggunakan Netbeans. Pada penelitian ini menggunakan data 823 record. Pertama kali Algoritma Genetika digunakan untuk mencari nilai K optimal yang nantinya akan digunakan di dalam Algoritma MKNN. Dari hasil percobaan, didapatkan nilai yang paling optimal adalah nilai $k = 2$. Selanjutnya klasifikasi dilakukan dengan Algoritma MKNN dengan menggunakan sebuah program yang dibuat dengan Netbeans. Dari hasil eksperimen yang dilakukan dengan menerapkan Algoritma Genetika pada Algoritma Modified K-Nearest Neighbour diperoleh hasil akurasi sebesar 96,47%, recall sebesar 96,47% dan precision sebesar 91,25%.

3.4 Komparasi Akurasi

Dari beberapa eksperimen yang telah dilakukan didapatkan hasil bahwa penggunaan Algoritma Genetika pada Algoritma Modified K-Nearest Neighbour diperoleh nilai yang optimal $k=2$, dan terbukti dapat meningkatkan akurasi untuk klasifikasi Pengajuan kredit.



Gambar 1 Komparasi Akurasi

Berdasarkan Gambar 1 dapat dilihat bahwa metode Algoritma Genetika dan MKNN memiliki hasil evaluasi akurasi yang lebih besar dan berarti memiliki akurasi yang lebih baik dibandingkan hasil akurasi metode menggunakan metode lain. Oleh karena itu, dapat disimpulkan bahwa metode Algoritma Genetika dan MKNN lebih baik dalam mendapatkan nilai evaluasi pada dataset pengajuan kredit.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Kesimpulan dari penelitian ini antara lain:

Dari hasil penelitian dan evaluasi peningkatan performa Algoritma Modified K-Nearest Neighbour dengan optimasi nilai K menggunakan Algoritma Genetika untuk klasifikasi pengajuan kredit pada koperasi simpan pinjam dapat disimpulkan beberapa hal bahwa dengan menerapkan optimasi nilai K menggunakan Algoritma Genetika, dapat meningkatkan performa Algoritma Modified K-Nearest Neighbour dalam mengklasifikasi nasabah yang berpotensi macet dan lancar. Performa Algoritma Modified K-Nearest Neighbour tergantung pada nilai K, dengan menentukan nilai K optimal menggunakan Algoritma genetika dapat meningkatkan performa Algoritma Modified K-Nearest Neighbour khususnya untuk nilai accuracy. Nilai accuracy yang diperoleh sebesar 96,47%, nilai recall sebesar 96,47% dan nilai precision sebesar 91,25%.

4.2 Saran

Hasil penelitian ini menunjukkan bahwa penggunaan Algoritma Genetika untuk optimasi nilai K dapat mempengaruhi performa Algoritma Modified K-Nearest Neighbour, sehingga disarankan untuk penelitian berikutnya Algoritma Genetika bisa digunakan untuk mengatasi masalah lain pada algoritma klasifikasi, misalkan Algoritma Genetika digunakan untuk seleksi atribut pada algoritma klasifikasi.

5. DAFTAR PUSTAKA

- Abdussomad and W. Gata, "ALGORITMA KLASIFIKASI C4.5 BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK PREDIKSI KELAYAKAN KREDIT," pp. 398–405, 2014.
- E. Arisawati, "Penerapan K-Nearest Neighbor Berbasis Genetic Algorithm Untuk penentuan Pemberian Kredit," *J. Sains Komput. Inform.*, no. 1, pp. 1–11, 2017.
- E. S. Y. Pandie, "Implementasi Algoritma Data Mining K-Nearest Neighbour (K-Nn) Dalam Pengambilan Keputusan Pengajuan Kredit," *Paradigma*, vol. X, no. Sainstek, pp. 31–34, 2012.
- H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J.*

- Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, no. 1, pp. 65–76, 2013.
- H. Parvin, H. Alizadeh, and B. Minaei-bidgoli, “MKNN : Modified K-Nearest Neighbor,” *Proc. World Congr. Eng. Comput. Sci. WCECS*, pp. 22–25, 2008.
- J. S. Informasi and F. Teknik, “Optimasi teknik klasifikasi modified k nearest neighbor menggunakan algoritma genetika,” *J. GAMMA*, no. September, pp. 130–134, 2014.
- J. Zurada and K. N. Kunene, “Comparisons of the performance of computational intelligence methods for loan granting decisions,” *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, pp. 1–10, 2011.
- Kusrini and L. E. Taufiq, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- N. Iriadi, “PENERAPAN ALGORITMA KLASIFIKASI DATA MINING DALAM,” vol. XIV, no. 2, pp. 120–129, 2012.
- N. Bhatia and C. Author, “Survey of Nearest Neighbor Techniques,” *IJCSIS) Int. J. Comput. Sci. Inf. Secur.*, vol. 8, no. 2, pp. 302–305, 2010.
- S. Susanto and D. Suryadi, *Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: Andi Offset, 2010.
- T. Faritcan and P. Siallagan, “Pencarian Nasabah dengan Menggunakan Data Mining dan Algoritma C 4 . 5 Koperasi Maduma Subang,” *J. Tek. Inform. dan Sist. Inf.*, vol. 1, pp. 221–228, 2015.