

OPTIMASI k-Nearst-Neighbor (k-NN) DENGAN PARTICLE SWARM OPTIMIZATION PADA KLASIFIKASI NASABAH KREDIT

Rachmat¹, Rudy Donny Liklikwati²

¹Politeknik Informatika Nasional (Polinas), ²STMIK Dipanegara Makassar
e-mail: ¹Rachmat27udinus@gmail.com, ²Rudy.donny@dipanegara.ac.id

Abstrak:

Data mining adalah teknik yang memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang sebelumnya tidak diketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting. Salah satu informasi yang dapat dihasilkan berupa pengetahuan dalam mengklasifikasi nasabah kredit. Sehingga pihak bank dapat memberikan kredit kepada nasabah, dalam mengklasifikasi nasabah kredit pihak bank dapat mengetahui bahwa nasabah ini lancar atau tidak lancar dalam membaya rangsuran. Data yang ada dianalisa menggunakan Particle Swarm Optimization pada algoritma k-Nearst-Neighbor. k-NN merupakan salah satu metode yang dapat menghilangkan variabel yang tidak signifikan dari model polynomial. Particle Swarm Optimization pada k-NN bertujuan untuk melakukan klasifikasi data nasabah kredit dengan akurasi 87,59% setelah digunakan optimization meningkat akurasi 90,73%, kemudian pola tersebut dapat digunakan oleh pihak perusahaan asuransi dalam pengambilan keputusan lancar atau tidak lancar nasabah dalam membayar angsuran.

Kata kunci: Data mining, *Particle Swarm Optimization (PSO)*, algoritma *k-Nearst-Neighbor (k-NN)*

Abstract

Data mining is a technique that utilizes large amounts of data to obtain valuable information that was previously unknown and can be used for important decision making. One of the information that can be generated is knowledge in classifying credit customers. So that the bank can provide credit to customers, in classifying credit customers the bank can know that this customer is smooth or not smooth in paying installments. Existing data were analyzed using Particle Swarm Optimization on the k-Nearst-Neighbor.k-NN algorithm is one method that can eliminate insignificant variables from the polynomial model. Particle Swarm Optimization in k-NN aims to classify credit customer data with an accuracy of 87.59% after optimization is used to improve its accuracy by 90.73%, then the pattern can be used by insurance companies in making current or non-current decisions in paying customers installments.

Keywords: Data mining, *Particle Swarm Optimization (PSO)*, *k-Nearst-Neighbor (k-NN) algorithm*.

1. PENDAHULUAN

Risiko kredit merupakan eksposur yang muncul akibat gagalnya pihak counterparty dalam memenuhi kewajibannya. Risiko ini muncul sebagai akibat satu atau lebih debitur yang memiliki kinerja buruk. Kinerja tersebut bisa berasal dari terpenuhinya seluruh atau sebagian

dari isi perjanjian kredit yang sudah disepakati secara bersama oleh pihak debitur. Hal inilah yang menjadi dasar dari perhatian pihak leasing dan bank yakni kondisi keuangan harga nilai pasar sebagai jaminan dan juga hal yang tak kalah penting yaitu karakter dari debitur.[1], Teknik klasifikasi ialah teknik pembelajaran yang dimanfaatkan untuk memprediksi nilai dari atribut kategori target. Klasifikasi bertujuan untuk mengelompokkan objek yang ditugaskan hanya ke salah satu nomor kategori yang disebut class.[3].

Klasifikasi nasabah kredit merupakan masalah yang menarik untuk diteliti. Beberapa riset bidang komputer untuk mengurangi klasifikasi nasabah kredit telah banyak dilakukan dalam rangka menghindari kerugian suatu perusahaan pembiayaan (leasing) atau bank. Dalam menyelesaikan masalah klasifikasi kredit nasabah, data akan dikelompokkan menjadi dua kategori yaitu kredit lancar dan tidak lancar. Jadi rasanya tepat menggunakan teknik klasifikasi pada data mining. Data mining merupakan sebuah proses yang bertujuan untuk memperoleh pola secara semi otomatis atau otomatis dari data yang sudah ada di dalam database yang digunakan untuk menyelesaikan suatu masalah.[2].

2. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian mengenai penerapan Algoritma k-Nearest Neighbor untuk penentuan risiko kredit kepemilikan kendaraan bermotor melalui teknik klasifikasi data mining telah banyak dilakukan. Henny Leidiyana melakukan penelitian di tahun 2013, dimana menggunakan dataset perusahaan pembiayaan (*leasing*) yang beralamat di cakarang. Paper ini membahas algoritma k-Nearest Neighbor (k-NN) yang diterapkan pada data nasabah yang menggunakan jasa keuangan kredit kendaraan bermotor. Hasil testing untuk mengukur performa algoritma ini menggunakan metode Cross Validation, Confusion Matrix dan kurva ROC dan menghasilkan akurasi dan nilai AUC berturut-turut 81,46 % dan 0,984. Karena nilai AUC berada dalam rentang 0,9 sampai 1,0 maka metode tersebut masuk dalam kategori sangat baik (excellent).[4].

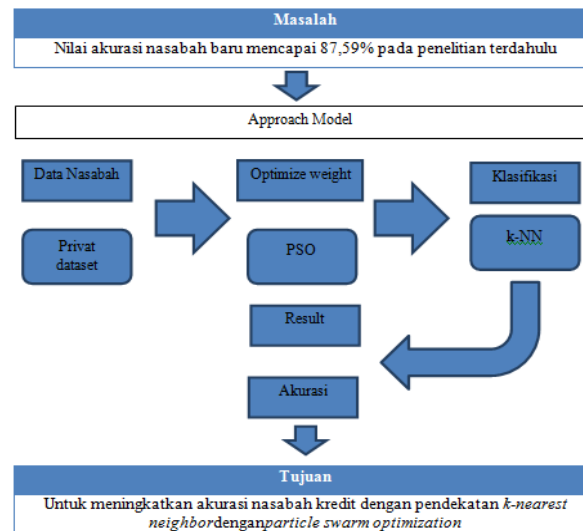
2.2 Landasan Teori

Proses atau tahapan dalam data mining dapat dibagi menjadi beberapa tahap, yaitu pembersihan data, integrasi data, transformasi data, aplikasi teknik data mining[10], evaluasi pola yang ditemukan, presentasi pengetahuan. Definisi lain dari data mining adalah kupulan dari beberapa proses yang pemodelan, mengeksplorasi dan memilih sejumlah besar data yang bertujuan untuk mengetahui pola data baru yang tidak diketahui sebelumnya untuk keuntungan bisnis [11].

2.3 Kerangka Pemikiran

Berdasarkan pertimbangan tersebut penggunaan *k-Nearest Neighbor* (k-NN) dengan pendekatan *particle swarm optimization*, diharapkan dapat meningkatkan performa dari

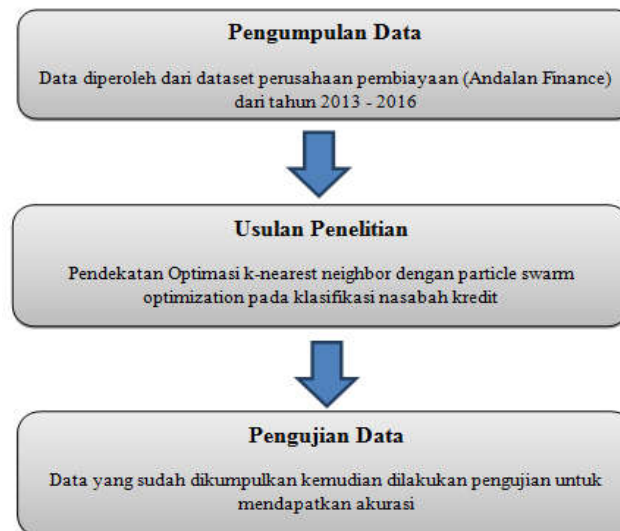
penentuan kelayakan kredit sehingga bisa mendapatkan nilai akurasi yang lebih tinggi. Gambar 1 berikut merupakan kerangka pemikiran dari penelitian ini:



Gambar 1. Kerangka pemikiran dari penelitian

3. METODE PENELITIAN

Adapun *research design* dari penelitian yang dilakukan adalah sebagai berikut:



Gambar 2. *Research design*

3.1 Jenis Penelitian

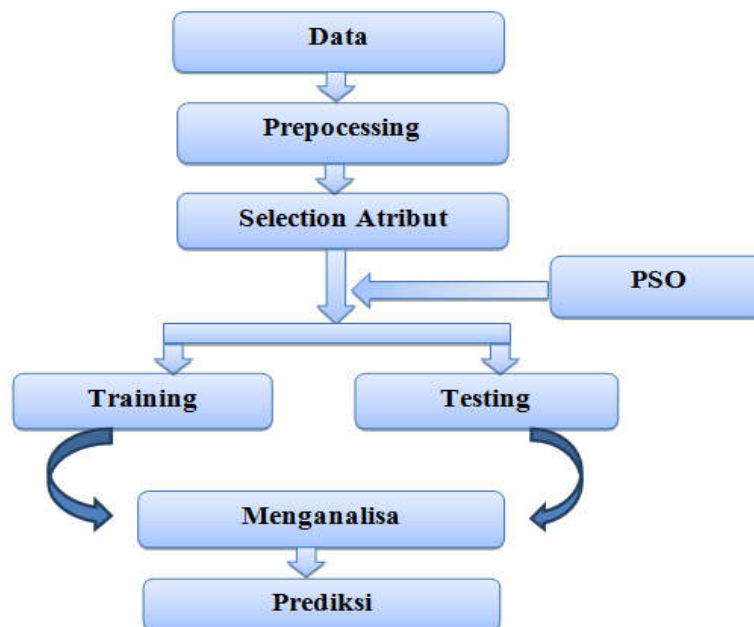
Jenis penelitian yang dilakukan adalah dengan metode eksperimen, yaitu untuk mengetahui pengaruh metode optimasi *Particle Swarm Optimization* pada K-nearest-neighbor dalam menghasilkan nilai akurasi yang lebih baik, dilakukan dengan sebuah data kemudian diimplementasikan pada *rapidminer* dan membandingkan pengujian data dengan tanpa menggunakan metode optimasi.

3.2 Metode Pengumpulan Data

Pengumpulan data yang digunakan dalam penelitian ini menggunakan data primer yang diperoleh dari data privat perusahaan pembiayaan yang bernama Andalan Finance yang terdiri dari 1244 record dan 44 atribut.

3.3 Metode Penyelesaian yang Diusulkan

Dalam melakukan prediksi terhadap hasil nasabah kredit diperlukan metode yang diusulkan untuk selanjutnya digunakan dalam prediksi nasabah kredit. Metode yang digunakan dalam penelitian ini adalah k-Nearest-Neighbor dengan bantuan optimasi *Particle Swarm Optimization (PSO)*. Metode yang diusulkan seperti pada Gambar 3. dibawah ini :



Gambar 3. Optimasi k-Nearest-Neighbor dengan PSO

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pengumpulan Data

Pada tahap pengumpulan data, dimana dataset yang digunakan pada penelitian didapat dari perusahaan pembiayaan di kota makassar dari tahun 2013 – 2016 yang berjumlah 1244 record dan 44 atribut. Berikut ini tabel 4.1 yang merupakan deskripsi dan pemetaan dari dataset Andalan Finance.

Tabel 1. Deskripsi dan pemetaan dari dataset Andalan Finance

| No. | Atribut | Tipe | Range | % | Keterangan |
|-----|-------------|------------|-------|-----|--------------------------|
| 1. | Regional | Polynomial | 1244 | 100 | B1 = 1244 = Regional 1 |
| 2. | Area | Polynomial | 1244 | 100 | C1 = 1244 = Area 4 |
| 3. | Nama Cabang | Polynomial | 1244 | 100 | D1 = 1244 = AFI Makassar |
| 4. | Tgl Setuju | Date_time | 372 | 37 | G1 = tahun 2016 |
| | | | 414 | 41 | G1 = tahun 2015 |
| | | | 348 | 34 | G1 = tahun 2014 |
| | | | 110 | 11 | G1 = tahun 2013 |

4.2 Eksperimen dan Pengujian Metode

Sebelum pengujian dilakukan terlebih dahulu mencari parameter yang sesuai sebagai atribut masukan untuk meningkatkan akurasi. Adapun parameter yang digunakan dalam pengujian terdiri dari *K*, *Population size*, *Inertia weight*, *local best weight*, *Global best weight*, *Min weight*, *Max weight* dan *Number of validation*. Hasil k-NN, PSO akan diuji dengan menggunakan *K-Fold Cross Validation*.

Tabel 2. Hasil Pengujian k-NN

| K | Population Size | Inertia weight | Local best weight | Global best weight | Min weight | Max weight | Number of validation | k-NN+ PSO Accuracy (%) |
|---|-----------------|----------------|-------------------|--------------------|------------|------------|----------------------|------------------------|
| 1 | 5 | 1.0 | 1.0 | 1.0 | 0.0 | 1.0 | 10 | 82.52% |
| 3 | 10 | 1.0 | 1.0 | 1.0 | 0.0 | 1.0 | 10 | 87.59% |
| 5 | 15 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 10 | 87.35% |
| 7 | 20 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 10 | 90.73% |

Hasil eksperimen k-NN menggunakan metode PSO diatas dengan $k = 7$, *population size* = 20, *inertia weight* = 1.0, *local best weight* = 1.0, *global best weight* = 1.0, *min weight* = 1.0, *max weight* = 1.0, *number of validation* = 10 menghasilkan nilai *accuracy* sebesar 90.73%.

Nilai Tabel 2. dapat diperoleh dengan menggunakan persamaan sebagai berikut:

$$\begin{aligned}
 \text{accuracy} &= \frac{a + d}{a + b + c + d} = \frac{TP + TN}{TP + FN + FP + TN} \\
 &= \frac{978 + 148}{978 + 97 + 18 + 148} = \frac{1126}{1241} = 0.90733 = 90.73\%
 \end{aligned}$$

$$\begin{aligned} \text{sensitivity} &= \frac{a}{a+b} = \frac{TP}{TP+FN} \\ &= \frac{978}{978+97} = \frac{978}{1075} = 0.90977 = 90.97\% \end{aligned}$$

$$\begin{aligned} \text{specificity} &= \frac{d}{c+d} = \frac{TN}{TN+FP} \\ &= \frac{148}{18+148} = \frac{148}{166} = 0.89167 = 89.16\% \end{aligned}$$

$$\begin{aligned} \text{PPV (Positive Predictive Value)} &= \frac{a}{a+c} = \frac{TP}{TP+FP} \\ &= \frac{978}{978+18} = \frac{978}{996} = 0.98193 = 98.19\% \end{aligned}$$

$$\begin{aligned} \text{NPV (Negative Predictive Value)} &= \frac{d}{b+d} = \frac{TN}{TN+FN} \\ &= \frac{148}{97+148} = \frac{148}{245} = 0.60408 = 60.41\% \end{aligned}$$

4.3 Pembahasan

Data hasil eksperimen di atas, baik pengujian dan evaluasi menggunakan kurva ROC dan *Confusion Matrix* terbukti bahwa model algoritma *k-Nearest Neighbor* (k-NN) pada *particle swarm optimization* (PSO) memiliki nilai akurasi yang lebih baik jika dibandingkan dengan model penelitian sebelumnya [4][5][6][7][8][9]. Adapun perbandingan akurasi dapat dilihat pada Tabel 3. di bawah ini.

Tabel 3. Tabel Perbandingan Akurasi

| Metode | Accuracy (%) |
|---------------------------------|--------------|
| k-Nearest Neighbor (k-NN) | 81.46 |
| SVM + Chi Square | 77.80 |
| k-Nearest Neighbor (k-NN) | 76.50 |
| k-Nearest Neighbor (k-NN) + PSO | 86.09 |
| k-Nearest Neighbor (k-NN) | 84.42 |
| k-Nearest Neighbor (k-NN) | 70.00 |
| k-Nearest Neighbor (k-NN) + PSO | 90.73 |

5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil eksperimen dan evaluasi penelitian yang dilakukan dengan dataset yang berbeda pada penelitian sebelumnya, maka dapat disimpulkan bahwa eksperimen dengan tingkat akurasi tertinggi pada metode *k-Nearest Neighbor (k-NN)* pada *particle swarm optimization* sebesar 90.73% dan nilai *Area Under Curve (AUC)* sebesar 0.464 berhasil memperbaiki akurasi penelitian sebelumnya yakni 86.09%.

5.2 Saran

Model *k-Nearest Neighbor (k-NN)* pada *particle swarm optimization (PSO)* telah diterapkan pada nasabah kredit. Berdasarkan hasil penelitian ini, penulis menyarankan bahwa masih besarnya kesempatan untuk penelitian di bidang klasifikasi nasabah kredit untuk meningkatkan nilai akurasi serta penggunaan data set yang berbeda sebagai pembanding akurasi yang dihasilkan untuk model *k-Nearest Neighbor (k-NN)* pada *particle swarm optimization (PSO)*.

DAFTAR PUSTAKA

- [1] A. Terhadap, P. Manajemen, R. Kredit, P. Pt, B. Ekspor, I. Dharma, S. Program, P. Sarjana, U. Gunadarma, A. K. Kunci, A. Terhadap, P. Manajemen, R. Kredit, B. E. Indonesia, M. R. Kredit, P. Latar, and B. Masalah, "No Title," pp. 1–19, 2007.
- [2] I. H. Witten and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations," 2000.
- [3] D. Hand, *Principles of Data Mining*, vol. 2001. 2001.
- [4] K. Kepemilikan and K. Bemotor, "Penerapan algoritma k-nearest neighbor untuk penentuan resiko kredit kepemilikan kendaraan bemotor," vol. 1, no. 1, pp. 65–76, 2013.
- [5] S. Santosa, P. Teknik, I. Universitas, and D. Nuswanto, "CREDIT SCORING MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DENGAN TEKNIK SELEKSI ATRIBUT BERBASIS CHI SQUARED STATISTIC DAN PARTICLE SWARM OPTIMIZATION," vol. 10, no. April, pp. 1–18, 2014.
- [6] A. Diskriminan, U. Analisis, R. Kredit, P. Koperasi, S. Pinjam, D. I. Kopinkra, and S. Rejeki, "Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 Yogyakarta, 15 November 2014 ISSN: 1979-911X," pp. 275–284, 2014.
- [7] Q. Cao and Y. Liu, "A KNN Classifier with PSO Feature Weight Learning Ensemble," pp. 110–114, 2010.
- [8] S. Ruggieri, "k-NN as an Implementation of Situation Testing for Discrimination Discovery and Prevention," pp. 502–510.
- [9] H. Kou, G. Gardarin, and B. L. A. Reine, "Study of Category Score Algorithms for k-NN Classifier," pp. 393–394.
- [10] S. Liao, P. Chu, and P. Hsiao, "Expert Systems with Applications Data mining

- techniques and applications – A decade review from 2000 to 2011,” *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11303–11311, 2012.
- [11] K. Tsipstis and A. Chorianopoulos, “Data Mining Techniques in CRM: Inside Customer Segmentation.”
- [12] M. Kamber, M. Kaufmann, and P. All, “Note : This manuscript is based on a forthcoming book by Jiawei Han Jiawei Han and Micheline Kamber,” 2000.
- [13] M. G. Epitropakis, “Business Intelligence : Optimization for Decision Making,” no. March, 2015.
- [14] W. Mansur and J. Polnar, “Microsoft SQL Server Installation Guide,” no. January, 2013.
- [15] J. A. Hoffer, M. B. Prescott, and F. R. Mcfadden, “Simpelt E-R diagram,” pp. 1–26, 2005.
- [16] D. King, J. E. Aronson, and P. Hall, “BUSINESS INTELLIGENCE University of Hawaii With contributions by.”
- [17] P. Bank, “NOMOR 10 TAHUN 1998 PERUBAHAN ATAS UNDANG-UNDANG NOMOR 7 TAHUN 1992 TENTANG PERBANKAN [LN 1998 / 82 , TLN 3790] tahun,” no. 3, 1998.
- [18] X. Li and Y. Zhong, “An Overview of Personal Credit Scoring : Techniques and Future Work,” vol. 2012, no. October, pp. 181–189, 2012.
- [19] H. Yulianton, “Data Mining untuk Dunia Bisnis Keputusan Informasi,” vol. XIII, no. 1, pp. 9–15, 2008.

Filename: JTRISTE Edisi 11
Directory: C:\Users\ajljftihh\Documents
Template: C:\Users\ajljftihh\AppData\Roaming\Microsoft\Templates\Normal.dotm
Title: Sebuah Kajian Pustaka:
Subject:
Author: cairo
Keywords:
Comments:
Creation Date: 4/13/2017 10:16:00 AM
Change Number: 240
Last Saved On: 4/20/2020 5:00:00 PM
Last Saved By: ajljftihh
Total Editing Time: 3,544 Minutes
Last Printed On: 4/20/2020 5:02:00 PM
As of Last Complete Printing
Number of Pages: 17 (approx.)
Number of Words: 17,959 (approx.)
Number of Characters: 102,368 (approx.)