

**CREDIT SCORING MENGGUNAKAN METODE
LOCAL MEANS BASED K HARMONIC NEAREST NEIGHBOR (MLMKHNN)**

Tatik Widiharih, Moch. Abdul Mukid

Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro, Indonesia

e-mail: mamukid@live.undip.ac.id

DOI: 10.14710/medstat.11.2.107-117

Article Info:

Received: 20 October 2018

Accepted: 28 February 2019

Available Online: 4 March 2019

Keywords:

Credit Scoring, Multi Local Means Based K Harmonic Nearest Neighbor, Local Means Based K Harmonic Nearest Neighbor

Abstract: Credit Scoring is designed so that lenders can easily make decisions regarding whether a loan proposal from a prospective customer is worthy of approval or not. This study examines the application of the Multi Local Means Based K Harmonic Nearest Neighbor (MLMKHNN) method in the case of motorcycle credit in a financial institution. The classification capability of this method in detecting potential borrowers into the credit category is either good or bad compared to its previous method, Local Means Based K Harmonic Nearest Neighbor (LMKNN). In this case the MLMKHNN method has not shown better performance than the LMKNN method. At the same level of total accuracy, MLMKHNN requires more numbers of neighbors than the number of neighbors required by the LMKNN method.

1. PENDAHULUAN

Kredit adalah sebuah fasilitas keuangan yang memungkinkan seseorang atau badan usaha untuk meminjam uang dan membayarnya kembali dalam jangka waktu yang ditentukan. Agar bisnis ini tidak menimbulkan permasalahan gagal bayar maka risikonya harus dikelola dengan baik dengan cara memprediksinya melalui metode tertentu. Prediksi terhadap risiko kredit ini dapat dilakukan melalui prosedur penilaian kredit (*credit scoring*). *Credit scoring* terkait dengan pengembangan model empiris untuk mendukung pengambilan keputusan dalam bisnis kredit (Crook *et al.*, 2007). Menurut Hand dan Henley (1997): “*Credit Scoring* adalah istilah yang digunakan untuk menggambarkan metode statistik formal yang digunakan untuk mengklasifikasikan pelamar kredit ke kelas risiko ‘baik’ dan ‘buruk’. Beberapa penelitian telah dilakukan untuk menguji keakuratan berbagai metode klasifikasi untuk penilaian kredit ini. Salah satu kajian perbandingan metode paling komprehensif hingga saat ini adalah studi yang dilakukan oleh Baesens *et al.* (2003). Salah satu metode klasifikasi yang sering digunakan untuk analisis penilaian kredit adalah metode *k nearest neighbor*.”

Aturan *k-Nearest Neighbor* (KNN) merupakan salah satu dari 10 algoritma teratas yang telah dipelajari secara ekstensif dan digunakan secara luas di bidang klasifikasi (Gou, *et al.* 2014). Aturan KNN bekerja sebagaimana sebuah teknik nonparametrik yang tidak

membutuhkan pengetahuan mengenai distribusi peluang dari peubah-peubah yang digunakan sebagai dasar untuk proses klasifikasi (Pan *et al.*, 2017). KNN mempunyai konsep dan implementasi yang sederhana. Selain daripada itu, KNN secara asimtotik memiliki kinerja yang optimal dalam kerangka Bayes yaitu ketika banyaknya sampel *training* (latih) N besar sedemikian sehingga rasio antara ukuran ketetanggaan k dan N mendekati nol. Oleh karena itu, pengklasifikasi yang didasarkan pada KNN sering kali mencapai kinerja klasifikasi yang baik pada beberapa aplikasi praktis.

Meskipun pengklasifikasi KNN yang standar mempunyai banyak kelebihan tetapi masih ada beberapa kekurangan. Salah satu isu yang sering muncul adalah sensitifitas terhadap pilihan ukuran ketetanggaan k . Jika k nilainya terlalu kecil, keputusan klasifikasi dapat menjadi salah karena sampel yang buruk dan adanya *noise*, sebaliknya nilai k yang terlalu besar dapat mengarah ke penurunan kinerja klasifikasi karena adanya pencilan dalam k tetangga terdekat yang datang dari kelas-kelas yang salah (Pan *et al.*, 2017). Faktanya kekurangan ini disebabkan adanya pencilan-pencilan yang berada di k tetangganya. Untuk mengatasi isu ini sebuah pilihan yang adaptif dari ukuran ketetanggaan yang didasarkan atas keyakinan statistik telah dikembangkan oleh Wang *et al.* (2006). Isu yang kedua adalah pendekatan untuk mengkombinasikan label-label kelas dalam k tetangga untuk membuat keputusan klasifikasi. Sebagaimana yang telah diketahui, suara mayoritas digunakan dalam KNN tetapi hal ini dapat menjadi sebuah masalah ketika tetangga-tetangga terdekat bervariasi di dalam jarak. Biasanya tetangga-tetangga yang lebih dekat lebih dapat diandalkan untuk memprediksi kelas dari sebuah amatan baru. Untuk mengatasi masalah ini, beberapa metode pembobotan untuk KNN telah dikembangkan dimana bobot yang lebih tinggi diberikan kepada tetangga-tetangga yang lebih dekat. Isu yang ketiga adalah bahwa aturan KNN tidak dapat mengklasifikasi secara tepat untuk amatan baru ketika data peubah-peubah yang digunakan mirip dengan data *training* pada kelas yang berbeda (Liu *et al.*, 2013). Karena seringkali sejumlah tetangga terdekat dari kelas yang berbeda mirip dengan k tetangga-tetangga terdekat dari amatan baru. Dalam hal ini aturan KNN akan salah dalam memberikan kelas terhadap amatan-amatan baru tersebut. Oleh karena itu, beberapa metode klasifikasi *fuzzy* telah dikembangkan dalam rangka untuk mengijinkan amatan-amatan baru berada pada kelas yang berbeda dengan tingkat keyakinan tertentu (Keller, *et al.* 1985).

Untuk merancang pengklasifikasi yang lebih *robust* terhadap pencilan, pengklasifikasi nonparametrik sederhana yang diberi nama *Local Mean k Nearest Neighbor* (LMKNN) telah diusulkan oleh Mitani dan Hamamoto (2006). Karena keefektifan dan desain klasifikasi LMKNN yang mudah dimana gagasan intinya telah berhasil diterapkan pada banyak metode perbaikan lainnya (Gou *et al.*, 2014; Samsudin, and Bradley, 2010; Zeng *et al.*, 2009). Pengelompokan LMKNN pertama kali diperkenalkan untuk mengatasi masalah ini dan meningkatkan kinerja klasifikasi. Dalam aturan LMKNN, vektor rata-rata lokal dihitung terlebih dahulu sesuai dengan semua tetangga terdekat k di masing-masing kelas, kemudian amatan baru ditempatkan ke kelas dengan jarak Euclid yang minimum di antara amatan tersebut dan vektor rata-rata lokal. Namun, pengelompokan LMKNN memilih nilai tunggal k untuk setiap kelas dan nilai seragam k untuk semua kelas, yang dapat menyebabkan kesalahan klasifikasi.

Untuk lebih meningkatkan kinerja klasifikasi LMKNN, beberapa metode KNN yang didasarkan atas rata-rata lokal telah diusulkan. Metode *pseudo NN* (PNN) dikembangkan untuk mengatasi masalah yang disebabkan oleh pilihan nilai tunggal k (Zeng *et al.*, 2009). Jumlah jarak tertimbang antara amatan baru dan tetangga terdekatnya di masing-masing kelas digunakan sebagai ukuran kesamaan antara amatan baru dan

pseudo sampel *training*-nya. Berdasarkan gagasan dasar pengklasifikasi LMKNN dan pengklasifikasi PKNN, *Local Mean Pseudo K Nearest Neighbor* (LMPKNN) diusulkan di Gou *et al.* (2014) dan menunjukkan kinerja klasifikasi yang menjanjikan. Pada klasifikasi LMPKNN, jarak *pseudo* diperoleh dengan menggabungkan jarak tertimbang antara amatan baru dan masing-masing vektor rata-rata lokal di kelas yang berbeda.

Pendekatan yang lain juga dikembangkan dalam rangka untuk menangani masalah yang disebabkan oleh adanya pencilan. *Mutual k Nearest Neighbor* (MKNN) (Liu, dan Zhang, 2012) mewakili sebuah teknik yang sederhana dan ampuh untuk menemukan tetangga terdekat yang lebih andal melalui prosedur penghilangan data. Selain itu Pan *et al.* (2017) juga mengusulkan metode *Multi Local Means Based k Harmonic Nearest Neighbor* (MLMKHNN) yang diklaim mampu mengatasi secara bersamaan permasalahan-permasalahan utama dalam KNN. Dalam metode ini mereka tidak menggunakan jarak Euclid sebagaimana biasa digunakan dalam metode KNN tetapi mereka menggunakan jarak harmonik yang diharapkan lebih *robust* terhadap pencilan.

Penelitian ini mengkaji kinerja klasifikasi metode *Multi Local Means Based k Harmonic Nearest Neighbor*. Untuk melihat apakah hasil yang diperoleh layak atau tidak, maka kinerja MLMKHNN dibandingkan dengan metode LMKNN yang merupakan metode yang sudah ada sebelumnya. Metode-metode tersebut dievaluasi dengan menggunakan data kredit dari sebuah lembaga keuangan di Kota Semarang, Propinsi Jawa Tengah yang menyalurkan kredit untuk pembelian sepeda motor. Penelitian mengenai penerapan metode MLM-KHNN untuk penilaian kredit belum ada yang melakukan, oleh karena itu penelitian ini diharapkan mampu menambah wawasan bagi akademisi dan analis kredit dalam pengembangan metode-metode penilaian kredit. Metode-metode klasifikasi nonparametrik seperti di atas menjadi salah satu alternatif pilihan manakala asumsi-asumsi pada pendekatan parametrik tidak terpenuhi.

2. TINJAUAN PUSTAKA

Pada bagian ini akan dijelaskan secara ringkas mengenai metode *k nearest neighbor* dan metode-metode utama dari penelitian ini yaitu meliputi LMKNN dan MLM-KHNN.

2.1. *K Nearest Neighbor*

Metode *Nearest Neighbor* adalah pendekatan nonparametrik pada kasus klasifikasi yang diusulkan oleh Fix dan Hodges (1952). Metode ini pertama kali digunakan untuk *credit scoring* oleh Chatterjee dan Barcun (1970) kemudian oleh Henley dan Hand (1996). Gagasan utama metode ini adalah menentukan jarak data pada ruang metrik untuk mengukur seberapa jauh lokasi antar data. Kemudian dengan sampel data yang lalu sebagai patokan, data baru diklasifikasikan ke dalam label kelas bergantung pada proporsi jumlah-*k* data terdekat dari sampel yang mewakili, yaitu tetangga terdekat (*nearest neighbors*) data baru (Thomas *et al.*, 2002).

Metode *Nearest Neighbor* termasuk dalam klasifikasi *lazy learner* karena menunda proses pelatihan (atau tidak melakukan sama sekali) sampai ada data *testing* yang ingin diketahui label kelasnya, maka metode baru akan menjalankan algoritmanya. Algoritma *Nearest Neighbor* melakukan klasifikasi berdasarkan kemiripan suatu data dengan data yang lain (Tan *et al.*, 2005 dalam Prasetyo, 2014).

Kemiripan antara data *testing* terhadap data *training* dapat diketahui secara geometris, semakin dekat lokasi data *training* terhadap data *testing*, maka bisa dikatakan

bahwa data *training* tersebut yang lebih dipandang mirip oleh data *testing*. Semakin dekat maka semakin miriplah data *testing* terhadap sejumlah k tetangga data *training* terdekat. Ukuran ketidakmiripan yang paling umum digunakan adalah jarak *Euclid* yang diformulasikan oleh persamaan berikut:

$$d_2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^r (x_i - y_i)^2}$$

Dengan $\mathbf{x}, \mathbf{y} \in X$ dan x_i, y_i adalah nilai fitur ke- i dari \mathbf{x} dan \mathbf{y} , sedangkan r adalah jumlah fitur dalam vektor. Ukuran ketidakmiripan *Euclid* akan memberikan $d_0 = 0$, maka jarak minimal yang mungkin diantara dua vektor data juga 0. Selain itu, jarak dari \mathbf{x} ke \mathbf{y} akan sama dengan \mathbf{y} ke \mathbf{x} , $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.

Pada algoritma KNN dengan data berdimensi q , jarak dari suatu data ke data yang lain dapat dihitung. Nilai jarak inilah yang digunakan sebagai nilai kedekatan/kemiripan antara data *testing* dengan data *training*. Nilai k pada KNN berarti k -data tetangga dari data *testing*. Jika k bernilai 1, kelas dari satu data *training* sebagai tetangga terdekat (terdekat pertama) dari data *testing* tersebut akan diberikan sebagai kelas untuk data *testing*. Begitu juga jika nilai k adalah 2, 3, 4, 5, dan sebagainya. Jika dalam k -tetangga ada dua kelas yang berbeda, akan diambil kelas dengan jumlah data terbanyak (*voting* mayoritas).

Ada beberapa isu kunci yang mempengaruhi kinerja KNN, diantaranya adalah pemilihan nilai k (Wu and Kumar, 2009). Jika k terlalu kecil maka berakibat hasil prediksi yang didapat bisa sensitif terhadap keberadaan *noise*. Di sisi lain, jika k terlalu besar maka tetangga terdekat terpilih mungkin terlalu banyak dari kelas lain yang sebenarnya tidak relevan karena jarak yang terlalu jauh. Untuk memperkirakan nilai k yang terbaik, bisa dilakukan menggunakan teknik validasi silang. Penting untuk menegaskan bahwa $k=1$ mungkin dapat mewakili nilai k yang lain, khususnya untuk himpunan data yang kecil, termasuk biasanya jika digunakan untuk penelitian atau latihan di kelas. Bagaimanapun juga, jika diberikan sampel yang cukup, maka nilai k yang lebih besar akan lebih tahan terhadap *noise* (Prasetyo, 2014).

2.2. Local Mean k -Nearest Neighbor (LMKNN)

Menurut Mitani dan Hamamoto (2006), LMKNN merupakan perluasan dari KNN. Tujuan LMKNN adalah untuk mengatasi efek negatif dari munculnya penciran dari himpunan data *training*. Prinsip dasar LMKNN adalah vektor rata-rata lokal dari k tetangga terdekat di setiap kelas digunakan untuk mengklasifikasikan setiap amatan baru. Misalkan $T = \{\mathbf{x}_i \in \mathbb{R}^m\}_{i=1}^n$ merupakan data *training* dari m atribut, dimana n adalah banyaknya data *training* dan $y_i \in \{c_1, c_2, \dots, c_m\}$ merupakan label kelas dari \mathbf{x}_i . $T_i = \{\mathbf{x}_{ij} \in \mathbb{R}^m\}_{j=1}^{n_i}$ merupakan himpunan bagian pada T dari kelas c_i . Menurut Gou, *et al.* (2012), algoritma LMKNN dirumuskan sebagai berikut:

Langkah 1: Mencari nilai k tetangga terdekat dari himpunan T_i dari setiap kelas c_i untuk setiap nilai \mathbf{x} . Misalkan $T_{ik}^{nn}(\mathbf{x}) = \{\mathbf{x}_{ij}^{nn} \in \mathbb{R}^m\}_{j=1}^k$ merupakan himpunan KNN dari \mathbf{x} pada kelas c_i menggunakan matriks jarak Euclid, dengan nilai $k \leq n_i$

Langkah 2: Menghitung vektor rata-rata lokal \mathbf{u}_{ik}^{nn} dari kelas c_i menggunakan himpunan $T_{ik}^{nn}(\mathbf{x})$,

$$\mathbf{u}_{ik}^{nn} = \frac{1}{k} \sum_{j=1}^k \mathbf{x}_{ij}^{nn}$$

Langkah 3: Mengalokasikan \mathbf{x} ke kelas c_i jika jarak diantara vektor rata-rata lokal untuk c_i dari amatan baru data dalam jarak Euclid yang minimum

$$c = \operatorname{argmin}_{c_i} (\mathbf{x} - \mathbf{u}_{ik}^{nn})^T (\mathbf{x} - \mathbf{u}_{ik}^{nn})$$

2.3. Multi Local Means Based k Harmonic Nearest Neighbor (MLMKHNN)

Pada bagian ini akan dijelaskan secara ringkas mengenai MLMKHNN. Metode ini dikembangkan untuk memperbaiki kinerja klasifikasi LMKNN dan mengurangi sensitivitas terhadap nilai tunggal dari ukuran ketetanggaan k di masing-masing kelas dan nilai k yang seragam untuk semua kelas. Misalkan $T = \{(\mathbf{x}_i, c_i)\}_{i=1}^n$ adalah sebuah sampel training dari M kelas dimana $\mathbf{x}_i \in \mathbb{R}^D$ dan c_i adalah label kelas yang bersesuaian dengan \mathbf{x}_i dan $c_i \in \{\omega_1, \omega_2, \dots, \omega_M\}$. Untuk kelas ω_j , misalkan $T_j = \{(\mathbf{x}_{i,j}^j, c_{i,j}^j)\}_{i=1}^{n_j}$ adalah sampel training dari kelas ω_j dengan ukuran sampel n_j . Dalam aturan MLM-KHNN label kelas ω_j dari sebuah amatan $\mathbf{x} \in \mathbb{R}^D$ diperoleh dengan melalui langkah-langkah sebagai berikut:

langkah 1: Untuk setiap kelas ω_j , temukan k tetangga terdekat dari \mathbf{x} yang dinotasikan

dengan $NN_{\omega_j}^k = \{(\mathbf{x}_{i,j}^{NN}, c_{i,j}^{NN})\}_{i=1}^k$ dari T_j dimana telah diurutkan dalam urutan naik berdasarkan jarak Euclid terhadap \mathbf{x} .

langkah 2: Hitung vektor rata-rata k multi lokal yang didasarkan atas r tetangga terdekat ($1 \leq r \leq k$) dari T_j di setiap kelas ω_j . Misalkan $m_{\omega_j}^k(\mathbf{x}) = \left\{ \bar{m}_{\omega_j}^r \right\}_{r=1}^k$

menotasikan vektor rata-rata k multi lokal dengan $\bar{m}_{\omega_j}^r = \frac{1}{r} \sum_{i=1}^r \mathbf{x}_{i,j}^{NN}$.

langkah 3: untuk setiap kelas ω_j , hitung jarak rata-rata harmonic antara \mathbf{x} dan k vektor

rata-rata multi lokal $m_{\omega_j}^k(\mathbf{x}) = \left\{ \bar{m}_{\omega_j}^r \right\}_{r=1}^k$ yang diperoleh dari langkah 2, yaitu

$$HMD(\mathbf{x}, m_{\omega_j}^k(\mathbf{x})) = \frac{k}{\sum_{r=1}^k \frac{1}{d(\mathbf{x}, \bar{m}_{\omega_j}^r)}}$$

langkah 4: tempatkan \mathbf{x} pada kelas ω_c yang mempunyai jarak rata-rata harmonik minimum,

$$\text{yaitu } \omega_c = \operatorname{argmin}_{\omega_j} \left[HMD(\mathbf{x}, m_{\omega_j}^k(\mathbf{x})) \right]$$

2.4. Kriteria Evaluasi

Matriks konfusi adalah alat yang berguna untuk menganalisis seberapa baik metode klasifikasi dapat mengenali data dari berbagai kelas. Sebuah matriks konfusi untuk m kelas adalah sebuah tabel paling tidak berukuran $m \times m$. Untuk pengklasifikasi yang memiliki akurasi yang baik, akan diwakili nilai-nilai disepanjang diagonal dari matriks konfusi tersebut. Dari matriks konfusi dapat dihitung nilai-nilai akurasi yang meliputi:

a. *Apparen Error Rate* (APER)

APER adalah nilai yang digunakan untuk melihat kemungkinan kesalahan dalam mengklasifikasikan objek. Nilai APER menyatakan proporsi sampel salah diklasifikasikan. Metode klasifikasi yang terbaik jika memiliki nilai APER terkecil sehingga.

$$\text{APER} = \frac{\text{Jumlah amatan yang salah klasifikasi}}{\text{total amatan}}$$

$$\text{Total Akurasi} = 1 - \text{APER}$$

b. Spesifisitas dan Sensitivitas

Spesifisitas adalah ukuran akurasi yang menyatakan seberapa banyak peminjam lancar oleh model diklasifikasikan sebagai peminjam yang lancar pula, sedangkan sensitivitas adalah ukuran yang menyatakan seberapa banyak peminjam macet oleh model klasifikasi diprediksi sebagai peminjam macet. Spesifitas dan sensitivitas didefinisikan sebagai berikut:

$$\text{Spesifisitas (\%)} = \frac{t_{\text{pos}}}{\text{pos}}$$

$$\text{Sensitivitas (\%)} = \frac{t_{\text{neg}}}{\text{neg}}$$

t_{pos} adalah jumlah peminjam lancar yang diklasifikasikan sebagai lancar, pos adalah jumlah peminjam yang lancar, t_{neg} adalah jumlah peminjam macet yang diklasifikasikan sebagai macet dan neg adalah jumlah peminjam yang lancar

3. METODE PENELITIAN

Data yang digunakan dalam makalah ini adalah debitur kredit sepeda motor di sebuah lembaga pembiayaan di Kota Semarang, Propinsi Jawa Tengah dari Januari 2016-Juli 2017 dengan jumlah data sebanyak 4420 data. Data ini terdiri dari dua kelas status yaitu peminjam yang baik (kelas 1) meliputi 4240 data dan peminjam yang macet (kelas 2) yang terdiri atas 180 data. Dalam penelitian ini, data dibagi menjadi data latih sebanyak 80% data uji sebanyak 20%. Proporsi ini diimplementasikan di setiap kelas, 80% data pelatihan dan 20% data pengujian di kelas 1, dan melakukan hal yang sama di kelas 2.

Prediksi status kredit dari calon peminjam, dilakukan dengan menggunakan karakteristik data historis dari peminjam yang terdiri atas beberapa variabel yaitu total pinjaman dengan satuan rupiah, pendapatan dengan satuan rupiah, pengeluaran dengan satuan rupiah, lama bekerja dengan satuan tahun, pembayaran angsuran dengan satuan rupiah, dan bunga kredit dengan satuan persen. Kemudian untuk mengatasi perbedaan satuan dari variabel-variabel yang digunakan, dalam penelitian ini digunakan standarisasi dengan formula

$$Z = \frac{X - \bar{X}}{SD_{\bar{X}}}$$

Salah satu parameter dalam algoritma metode klasifikasi KNN adalah nilai k yang merupakan ukuran seberapa banyak amatan terdekat yang menjadi dasar dalam keputusan klasifikasi. Untuk memilih nilai k , ditentukan secara eksperimental. Dimulai dengan $k = 1$, kami menggunakan satu himpunan data uji untuk memperkirakan tingkat kesalahan dari klasifikasi. Proses ini dapat diulang pada nilai k yang berbeda. Nilai k yang memberikan

tingkat kesalahan minimum dan akurasi tertinggi dapat dipilih. Kami mencoba dalam beberapa nilai k dengan nomor ganjil agar sesuai dengan aturan mayoritas berdasarkan k-NN dimulai dengan $k = 1$ sampai $k = 25$ tergantung pada jumlah dataset. Setelah nilai k ditetapkan, kemudian dilakukan proses prediksi status kredit dari data uji. Selanjutnya nilai akurasinya dihitung. Dalam hal ini digunakan ukuran sensitivitas, spesifisitas, dan total akurasi.

3. HASIL DAN PEMBAHASAN

Saat ini, karena terus meningkatnya permintaan di pasar kredit sepeda motor, lembaga keuangan menghadapi lebih banyak tantangan daripada sebelumnya. Di masa lalu, sebagian besar manajer penilaian kredit menilai pelamar melalui indikator-indikator yang terlihat kemudian dikombinasikan dengan pengalaman mereka sendiri. Namun, pola konsumen dengan kredit negatif tidak selalu dapat dirangkum dengan indikator-indikator yang eksplisit dan penilaian berdasarkan pengalaman menghasilkan kegagalan konsumen dan kerugian bagi lembaga keuangan. Akibatnya, pendekatan melalui metode-metode kuantitatif telah diterapkan dalam penilaian kredit, sehingga pola intrinsik konsumen dapat ditemukan, dan prediksi kredit pelamar baru dapat dibuat. Penentuan kredit pemohon oleh model klasifikasi menghasilkan akurasi yang lebih tinggi daripada intuisi manusia. Penelitian ini mengkaji kinerja MLM-KHNN untuk penilaian kredit di sebuah lembaga keuangan yang menyalurkan kredit untuk pembelian sepeda motor.

Tabel 1 merupakan ringkasan statistik dari variabel-variabel karakteristik debitur tahun 2016 dan 2017 dari sebuah lembaga keuangan yang menyalurkan kredit untuk pembelian sepeda motor yang terdiri atas debitur yang lancar maupun yang macet dalam pembayaran kreditnya. Dengan menggunakan uji Mann Whitney, pada tingkat kepercayaan 90%, median dari jumlah pinjaman dari debitur lancar dan macet berbeda secara nyata. Hal ini menunjukkan bahwa variabel jumlah pinjaman dapat dijadikan dasar untuk mengklasifikasikan ke status mana calon peminjam kredit baru.

Variabel-variabel lain yang meliputi pendapatan bersih, pengeluaran, lama bekerja, dan besarnya angsuran tiap bulan menunjukkan perilaku yang sama dengan variabel besar pinjaman dimana rata-rata di setiap variabel dari peminjam yang lancar lebih besar daripada peminjam yang macet. Namun demikian untuk variabel bunga flat, rata-rata bunga dari peminjam yang lancar lebih rendah dari mereka yang macet. Untungnya, pada tingkat kepercayaan 90%, median dari tiap variabel pada peminjam lancar dan macet berbeda secara nyata. Oleh karena itu variabel-variabel tersebut dapat digunakan sebagai dasar untuk mengidentifikasi potensi bayar dari calon peminjam kredit.

Tabel 2 memperlihatkan nilai akurasi metode LMKNN untuk analisis penilaian kredit. Ukuran akurasi yang digunakan meliputi, sensitivitas, spesifisitas, dan total akurasi. Secara umum, nilai sensitivitas dari metode ini berfluktuasi dimana nilainya bergantung pada nilai k . Ketika $k = 1$ nilai sensitivitasnya adalah 0,726 dan nilainya akan terus naik sampai dengan $k = 15$ yang nilai sensitivitasnya mencapai 1. Namun setelah $k = 15$, nilai sensitivitas metode ini turun. Hal yang sama terjadi pada nilai spesifisitas metode LMKNN yang cenderung naik meskipun dalam kondisi fluktuatif. Untuk $k = 1$ nilai spesifisitas adalah 0,650 dan nilainya akan terus naik sampai dengan $k = 5$ yang mencapai 1. Namun demikian pada $k = 7$ nilai spesifisitasnya turun menjadi 0,972. Untungnya untuk nilai-nilai k selanjutnya nilainya sudah mencapai maksimum, yaitu 1.

Tabel 1 Statistik Deskriptif dari Variabel-Variabel Penelitian

| Statistik | | Minimum | Maksimum | Rata-rata | Median |
|-------------------------------|--------|-----------|-------------|--------------|---------------|
| Jumlah Pinjaman (Rupiah) | Lancar | 1.500.000 | 85.000.000 | 9.495.647,66 | 8.720.000,00 |
| | Macet | 2.130.000 | 23.973.824 | 8.416.118,52 | 7.270.000,00 |
| p-value | | | | | 0,0257 |
| Pendapatan Bersih (Rupiah) | Lancar | 1.700.000 | 120.000.000 | 5.241.538,92 | 4.000.000,00 |
| | Macet | 1.800.000 | 30.000.000 | 4.901.627,78 | 3.950.000,00 |
| p-value | | | | | 0,0615 |
| Pengeluaran (Pengeluaran) | Lancar | 1.080.000 | 30.000.000 | 2.620.974,06 | 2.000.000,00 |
| | Macet | 1.200.000 | 20.000.000 | 2.586.666,67 | 2.000.000,00 |
| p-value | | | | | 0,0267 |
| Lama Bekerja (Tahun) | Lancar | 2 | 50 | 10,40 | 8 |
| | Macet | 1 | 35 | 9,02 | 6 |
| p-value | | | | | 0,0139 |
| Angsuran (Rupiah) | Lancar | 140.000 | 7.000.000 | 633.742,81 | 595.000 |
| | Macet | 220.000 | 1.860.000 | 580.483,33 | 570.000 |
| p-value | | | | | 0,0099 |
| Bunga (Persen) | Lancar | 4,90 | 46,32 | 21,83 | 20,33 |
| | Macet | 14,99 | 37,50 | 22,60 | 21,39 |
| p-value | | | | | 0,0178 |

Tabel 2 Nilai Akurasi Metode LMKNN untuk Penilaian Kredit

| k | Sensitivitas | Spesifisitas | Total Akurasi |
|-----------|--------------|--------------|---------------|
| 1 | 0,726 | 0,250 | 0,707 |
| 3 | 0,978 | 0,750 | 0,938 |
| 5 | 0,983 | 1,000 | 0,945 |
| 7 | 0,993 | 0,972 | 0,954 |
| 9 | 0,995 | 1,000 | 0,955 |
| 11 | 0,999 | 1,000 | 0,958 |
| 13 | 0,998 | 1,000 | 0,957 |
| 15 | 1,000 | 1,000 | 1,000 |
| 17 | 1,000 | 1,000 | 1,000 |
| 19 | 1,000 | 1,000 | 1,000 |
| 21 | 0,998 | 1,000 | 0,957 |
| 23 | 0,998 | 1,000 | 0,957 |
| 25 | 0,996 | 1,000 | 0,956 |

Untuk menentukan parameter k terbaik pada metode LMKNN, digunakan ukuran akurasi total sebagai dasarnya. Metode ini memiliki kinerja terbaik pada $k = 15$ dengan akurasi total mencapai 100%. Ini berarti bahwa metode LMKNN dapat mengklasifikasikan dengan benar peminjam yang lancar dan peminjam macet sebesar 100%. Pada $k = 15$, nilai spesifisitas sebesar 1 yang berarti bahwa metode LMKNN ketika diterapkan pada kasus kredit sepeda motor di Indonesia menunjukkan jumlah peminjam lancar yang oleh model diklasifikasikan sebagai peminjam lancar adalah 100%. Pada nilai yang sama dari k , nilai

sensitivitasnya adalah 1 yang memiliki interpretasi bahwa ada 100% debitur macet oleh metode LMKNN diklasifikasikan sebagai debitur-debitur yang macet.

Tabel 3 menunjukkan nilai akurasi metode MLMKHNN untuk analisis penilaian kredit. Ukuran akurasi yang digunakan meliputi, sensitivitas, spesifisitas, dan total akurasi. Secara umum, nilai sensitivitas dari metode ini monoton naik seiring dengan bertambahnya nilai k . Ketika $k = 1$ nilai sensitivitasnya adalah 0,968 dan nilainya akan terus naik hingga ketika $k = 23$ nilai sensitivitasnya telah mencapai 1. Pada ukuran spesifitas, metode MLMKHNN memiliki nilai yang seragam, yaitu 1.

Untuk menentukan parameter k terbaik pada metode MLMKHNN, digunakan ukuran akurasi total sebagai dasar. Metode ini memiliki kinerja terbaik pada $k = 23$ dengan akurasi total mencapai 1. Ini berarti bahwa metode MLMKHNN dapat mengklasifikasikan dengan benar peminjam yang lancaran peminjam macet sebesar 100%. Pada $k = 23$, nilai spesifisitas sebesar 1 yang berarti bahwa metode MLMKHNN ketika diterapkan pada kasus kredit sepeda motor di Indonesia menunjukkan jumlah peminjam lancar yang oleh model diklasifikasikan sebagai peminjam lancar adalah 100%. Pada nilai yang sama dari k , nilai sensitivitasnya adalah 1 yang memiliki interpretasi bahwa ada 100% debitur macet oleh metode MLMKHNN diklasifikasikan sebagai debitur-debitur yang macet.

Tabel 3 Nilai Akurasi Metode MLMKHNN untuk Penilaian Kredit

| k | Sensitivitas | Spesifisitas | Total Akurasi |
|-----------|---------------------|---------------------|----------------------|
| 1 | 0,967 | 1,000 | 0,927 |
| 3 | 0,971 | 1,000 | 0,931 |
| 5 | 0,983 | 1,000 | 0,943 |
| 7 | 0,989 | 1,000 | 0,949 |
| 9 | 0,995 | 1,000 | 0,955 |
| 11 | 0,996 | 1,000 | 0,956 |
| 13 | 0,998 | 1,000 | 0,957 |
| 15 | 0,999 | 1,000 | 0,958 |
| 17 | 0,999 | 1,000 | 0,958 |
| 19 | 0,999 | 1,000 | 0,958 |
| 21 | 0,999 | 1,000 | 0,958 |
| 23 | 1,000 | 1,000 | 1,000 |
| 25 | 1,000 | 1,000 | 1,000 |

Dari uraian di atas, metode MLMKHNN belum mampu memperbaiki kinerja klasifikasi dari metode LMKNN khususnya pada kasus kredit sepeda motor ini. Untuk mencapai total akurasi yang terbaik yaitu 100%, metode MLMKHNN membutuhkan ukuran ketetapan $k = 23$, sedangkan pada LMKNN hanya membutuhkan $k = 15$. Ukuran tetangga yang lebih sedikit akan mempercepat proses komputasi untuk prediksi potensi status kredit dari calon nasabah.

4. KESIMPULAN

Credit Scoring telah menjadialah satu tugas penting dalam manajemen risiko untuk setiap perusahaan di bidang keuangan. Di satu sisi ini adalah demi kepentingan untuk menghindari persetujuan kredit dari pelanggan yang mungkin gagal bayar. Di sisi lain, regulator memerlukan sistem manajemen risiko yang ketat untuk melindungi

pelanggan mereka dan untuk menghindari kebangkrutan dengan dampak negatif pada ekonomi secara keseluruhan.

MLMKHNN merupakan pengembangan dari metode LMKNN yang oleh penemunya diklaim mampu memperbaiki kinerja LMKNN terkait dengan isu-isu penting dari metode klasifikasi berbasis pada aturan tetangga terdekat. Namun demikian dalam kasus ini metode MLMKHNN belum menunjukkan kinerja yang lebih baik dari metode LMKNN. Pada tingkat akurasi yang sama, MLMKHNN membutuhkan banyaknya tetangga yang lebih banyak dari pada banyaknya tetangga yang dibutuhkan dalam metode LMKNN.

5. UCAPAN TERIMA KASIH

Terima kasih yang sebesar besarnya kami sampaikan kepada semua pihak yang telah membantu terlaksananya penelitian ini. Penelitian ini dibiayai oleh Kementerian Riset, Teknologi dan Pendidikan Tinggi berdasarkan kontrak nomor:101-113/UN7.P4.3/PP/2018.

DAFTAR PUSTAKA

- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., dan Vanthienen, J. 2003. Benchmarking State-of-the-Art Classification Algorithms for Creditscoring. *Journal of the Operational Research Society*, Vol 54 No 6, 627–635.
- Chatterjee, S. dan Barcun, S. 1970. A Nonparametric Approach to Credit Screening. *Journal of the American Statistical Association*, Vol 65 No 329, 150–154.
- Cover, T. dan Hart, P. 1967. Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, Vol 13, 21–27.
- Crook, J. N., Edelman, D. B., dan Thomas, L. C. 2007. Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, Vol 183 No 3, 1447–1465.
- Dudani, S. A. 1976. The Distance-Weighted k-Nearest Neighbor Rule. *IEEE Transaction on System, Man and Cybernetics*, Vol 6 No 4, 325-327.
- Fix, E. dan Hodges, J. L. 1951. Discriminatory Analysis Nonparametric Discrimination. Consistency Properties. *Technical Report TR4*, USAF School of Aviation Medicine, Randolph Field, TX.
- Gou, J., Yi, Z., Du L., dan Xiong T. 2012. A Local Mean-Based k-Nearest Centroid Neighbor Classifier. *The Computer Journal*, Vol. 55 No. 9. 1058-1071.
- Gou, J., Zhan, Y., Rao, Y., Shen, X., Wang, X., dan He, W. 2014. Improved Pseudo Nearest Neighbor Classification. *Knowledge-Based Systems*, 70 (C), 361–375.
- Hand, D. J. dan Henley, W. E. 1997. Statistical Classification Methods in Consumer Credit Scoring: a Review. *Journal of Royal Statistical Social*, Vol 3, 523-541.
- Hastie, T., Tibshirani R., dan Buja A. 1994. Flexible Discriminant Analysis by Optimal Scoring. *Journal of the American Statistical Association*, Vol 89, 1255-1270.
- Keller, J.M., Gray, M.R., dan Givens, J.A. 1985. *IEEE Transactions On Systems, Man, And Cybernetics*, Vol. SMC-15 No. 4.
- Liu, H. dan Zhang, S. 2012. Noisy Data Elimination Using Mutual K-Nearest Neighbor for Classification Mining. *Journal of Systems and Software*, Vol 85 No 5, 1067-1074.

- Liu, Z., Pan Q., dan Dezer J. 2013. A New Belief-Based K-Nearest Neighbor Classification Method. *Pattern Recognition*, Vol 46, 834-844.
- Mitani Y dan Hamamoto, Y. 2006. A Local Mean-Based Nonparametric Classifier. *Pattern Recognition Letters*, Vol 27, 1151-1159.
- Pan Z., Wang Y., dan Ku, W. 2017. A New K Harmonic Nearest Neighbor Classifier Based on the Multi Local Means. *Expert System with Applications*, Vol. 67, 115-125.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Penerbit ANDI, Yogyakarta.
- Sansudin, N. A. dan Bradley, A. P. 2010. Nearest Neighbour Group-Based Classification. *Pattern Recognition*, Vol 43 No 10, 3458–3467.
- Thomas, L. C., Edelman, D. B., dan Crook, L. N. 2002. *Credit Scoring and Its Applications*. Society for Industrial and Applied Mathematics, Philadelphia.
- Wang, J., Neskovic, P., dan Cooper, L. N. 2006. Neighborhood Size Selection in The K - Nearest-Neighbor Rule using Statistical Confidence. *Pattern Recognition*, Vol. 39 No. 3, 417–423.
- Wu, X. dan Kumar, V. 2009. *The Top Ten Algorithms in Data Mining*. CRC Press Taylor and Analitic Machine Intelligence, London.
- Zeng, Y., Yang, Y., dan Zhao, L., 2008. *Pseudo Nearest Neighbor Rule For Pattern Classification*. Department of Automation, Shanghai Jiao Tong University, Elsevier Ltd.