

## Analisis Kesehatan Bank Menggunakan *Local Mean K-Nearest Neighbor* dan *Multi Local Means K-Harmonic Nearest Neighbor*

Alwi Assegaf<sup>1</sup>, Moch. Abdul Mukid<sup>2</sup>, Abdul Hoyyi<sup>3</sup>

<sup>1,2,3</sup>Departemen Statistika FSM Universitas Diponegoro

[alwiassegaf@student.undip.ac.id](mailto:alwiassegaf@student.undip.ac.id)

### ABSTRACT

The classification method continues to develop in order to get more accurate classification results than before. The purpose of the research is comparing the two k-Nearest Neighbor (KNN) methods that have been developed, namely the Local Mean k-Nearest Neighbor (LMKNN) and Multi Local Means k-Harmonic Nearest Neighbor (MLM-KHNN) by taking a case study of listed bank financial statements and financial statements complete recorded at Bank Indonesia in 2017. LMKNN is a method that aims to improve classification performance and reduce the influence of outliers, and MLM-KHNN is a method that aims to reduce sensitivity to a single value. This study uses seven indicators to measure the soundness of a bank, including the Capital Adequacy Ratio, Non Performing Loans, Loan to Deposit Ratio, Return on Assets, Return on Equity, Net Interest Margin, and Operating Expenses on Operational Income with a classification of bank health status is very good (class 1), good (class 2), quite good (class 3) and poor (class 4). The measure of the accuracy of the classification results used is the Apparent Error Rate (APER). The best classification results of the LMKNN method are in the proportion of 80% training data and 20% test data with  $k=7$  which produces the smallest APER 0,0556 and an accuracy of 94,44%, while the best classification results of the MLM-KHNN method are in the proportion of 80% training data and 20% test data with  $k=3$  which produces the smallest APER 0,1667 and an accuracy of 83,33%. Based on APER calculation shows that the LMKNN method is better than MLM-KHNN in classifying the health status of banks in Indonesia.

**Keywords:** Classification, Local Mean k-Nearest Neighbor (LMKNN), Multi Local Means k-Harmonic Nearest Neighbor (MLM-KHNN), Measure of accuracy of classification

### 1. PENDAHULUAN

Perbankan merupakan salah satu komponen penting perekonomian setiap negara dalam berbagai aspek kehidupan. Peranan perbankan saat ini sangat dominan dalam sistem keuangan, sehingga pemahaman dan pengelolaan bank yang baik tentu akan mendorong sistem keuangan yang baik. Dalam hal ini, Bank Indonesia sebagai bank sentral bertanggung jawab untuk mengawasi dengan cara mengetahui bagaimana keadaan keuangan serta kegiatan usaha masing-masing bank. Bank Indonesia secara berkala mengadakan suatu standar pengawasan dengan melakukan penilaian terhadap tingkat kesehatan suatu bank untuk mengantisipasi gagalnya suatu bank dalam menyimpan dan mengelola uang nasabahnya. (Dendawijaya, 2005).

Santosa (2007) mengatakan kemajuan dalam teknologi akuisisi data yang berfungsi untuk mengambil, mengumpulkan dan menyiapkan data, hingga memprosesnya untuk menghasilkan data yang dikehendaki. Minat akan kepemilikan *database* itu terus berkembang seiring dengan kemajuan di bidang teknologi dan sains yang memungkinkan mendapatkan informasi yang mungkin bernilai. Disiplin ilmu yang terkait dengan hal ini dikenal sebagai *data mining*. Larose (2005) mengatakan *data mining* merupakan analisis pengamatan *database* untuk menemukan hubungan yang tidak terduga untuk meringkas data dengan cara baru yang dapat dimengerti dan bermanfaat kepada pemilik data.

Ada banyak metode klasifikasi dalam *data mining*. Salah satu metode klasifikasi yang terdapat dalam *data mining* yaitu *k-Nearest Neighbor* (KNN). KNN merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data

yang lain. Teknik ini termasuk dalam kelompok klasifikasi nonparametrik (Prasetyo, 2012).

Menurut Mitani dan Hamamoto (2006), metode *Local Mean k-Nearest Neighbor* (LMKNN) merupakan perluasan dari KNN. LMKNN adalah klasifikator yang bertujuan untuk meningkatkan kinerja klasifikasi dan juga mengurangi pengaruh *outlier* yang ada, terutama dalam situasi ukuran sampel pelatihan yang kecil. Prinsip dasarnya yaitu menggunakan vektor rata-rata lokal dari  $k$  tetangga terdekat di setiap kelas untuk mengklasifikasi setiap himpunan data dalam keputusan pengklasifikasian.

Metode *Multi Local Means k-Harmonic Nearest Neighbor* (MLM-KHNN) mendeskripsikan penggolongan tetangga terdekat multi lokal yang berbasis  $k$ -harmonik. Tujuan dari aturan MLM-KHNN adalah untuk meningkatkan kinerja klasifikasi dan mengurangi kepekaan terhadap nilai tunggal dari lingkungan berukuran  $k$  untuk setiap kelas dan nilai seragam  $k$  untuk semua kelas dalam aturan LMKNN. Jarak rerata harmonik diperkenalkan untuk mengukur jarak antara sepasang kelompok titik. Ide dasar jarak rata-rata harmonik adalah mengambil jarak *euclidean* jumlah rata-rata harmonik antara satu titik data yang diberikan dan setiap titik data dalam kelompok titik yang lain (Pan *et al.*, 2016).

## 2. TINJAUAN PUSTAKA

### 2.1 Uji Kruskal Wallis

Menurut Priyatno (2013), Uji Kruskal Wallis adalah uji nonparametrik berbasis peringkat yang tujuannya untuk menentukan adakah perbedaan signifikan secara statistik antara dua atau lebih kelompok variabel independen pada variabel dependen yang berskala data numerik (interval atau rasio) dan skala ordinal. Hipotesisnya yaitu:

$H_0$  : tidak ada pengaruh variabel bebas terhadap variabel terikat

$H_1$  : ada pengaruh variabel bebas terhadap variabel terikat

Dengan daerah keputusan sebagai berikut:

Jika nilai *chi-square* hitung  $> \chi^2_{(\alpha,df)}$  maka hipotesis  $H_0$  ditolak, jika *chi-square* hitung  $< \chi^2_{(\alpha,df)}$  maka  $H_0$  gagal ditolak

atau

Jika *p-value*  $<$  daerah kritis maka keputusan hipotesis adalah menolak  $H_0$  dan menerima  $H_1$  atau yang berarti ada pengaruh variabel bebas terhadap variabel terikat

Jika *p-value*  $>$  daerah kritis maka keputusan hipotesis adalah menerima  $H_0$  dan menolak  $H_1$  atau yang berarti tidak ada pengaruh variabel bebas terhadap variabel terikat

Untuk melihat nilai *chi-square* tabel, maka didasarkan pada derajat kebebasan (dk) atau *degree of freedom* (df) =  $k-1$ , dan taraf signifikansi ( $\alpha$ ). Adapun rumus Kruskal Wallis yang digunakan adalah sebagai berikut:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \quad (1)$$

Keterangan:  $N$  = jumlah sampel

$R_i$  = jumlah peringkat pada kelompok  $i$

$n_i$  = jumlah sampel pada kelompok  $i$

### 2.2 Jarak Mahalanobis

Sebagian besar set data berisi satu atau beberapa pengamatan tidak biasa yang tampaknya tidak termasuk dalam pola variabilitas yang dihasilkan oleh pengamatan lain. Dengan data pada satu karakteristik, pengamatan yang tidak biasa adalah pengamatan yang sangat besar atau relatif sangat kecil dibanding yang lain. *Outlier* paling baik dideteksi

secara visual bilamana memungkinkan (Johnson and Wichern, 2007). Rumus untuk mendeteksi *outlier* multivariat diberikan sebagai berikut:

$$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \quad (2)$$

dengan:

$$\mathbf{x}_j = x_{j1}, x_{j2}, \dots, x_{jk}$$

$$\bar{\mathbf{x}} = \bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$$

$\mathbf{S}^{-1}$  = *generalized inverse* matriks varian kovarian dari vektor  $\mathbf{x}_j$

Cara mengidentifikasi terjadinya *multivariate outliers* adalah dengan menggunakan *box plot* atau statistik  $d^2$  (*Mahalanobis Distance*) dan dibandingkan dengan nilai  $\chi^2$  dengan tingkat kesalahan  $\alpha$ , dan derajat bebas  $df$  sebanyak variabel yang dianalisis. Jika  $d^2 > \chi^2_{\alpha;df}$  maka terdapat *outlier*, jika  $d^2 < \chi^2_{\alpha;df}$  maka tidak terdapat *outlier*.

### 2.3 Jarak Euclid

Secara geometris, pengukuran kedekatan akan diukur berdasarkan jarak antara data yang pertama terhadap data yang kedua. Semakin dekat jarak kedua data, maka semakin besar kemiripannya dan semakin jauh jarak kedua data, maka semakin kecil kemiripannya (Prasetyo, 2012). Ada banyak model pengukuran jarak, dan yang paling sering digunakan adalah jarak *Euclid* (Bezdek, 1981). Jarak *Euclid* memberikan jarak lurus antara dua buah data  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  dan  $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jp})$ .

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2} \quad (3)$$

$d$  adalah jarak antara data  $\mathbf{x}_i$  dan  $\mathbf{x}_j$ ,  $\mathbf{x}_i$  dan  $p$  adalah banyaknya dimensi data.

### 2.4 Local Mean K-Nearest Neighbor (LMKNN)

Metode LMKNN merupakan perluasan dari metode KNN. Aturan  $k$  tetangga terdekat berbasis lokal (Mitani dan Hamamoto, 2006) adalah telah menunjukkan bahwa metode ini dapat meningkatkan kinerja klasifikasi dan juga mengurangi pengaruh *outlier* yang ada, terutama dalam situasi ukuran sampel pelatihan kecil. Prinsip dasarnya yaitu menggunakan vektor rata-rata lokal dari  $k$  tetangga terdekat di setiap kelas untuk mengklasifikasikan setiap himpunan data dalam keputusan pengklasifikasian.

Dalam aturan LMKNN, diberikan sampel himpunan  $\mathbf{x} \in R^D$  yang diklasifikasikan ke dalam kelas  $\omega_c$  dengan langkah sebagai berikut:

Langkah 1. Temukan  $k$  tetangga terdekat untuk  $\mathbf{x}$  dari  $Tr_j$  untuk masing-masing kelas  $\omega_j$ . Misalkan  $NN_{\omega_j}^k(\mathbf{x}) = \{(y_{i,j}^{NN}, c_{i,j}^{NN})\}_{i=1}^k$  menunjukkan himpunan  $k$  tetangga terdekat dari sampel himpunan  $x$  di kelas  $\omega_j$ , di mana  $(y_{i,j}^{NN}, c_{i,j}^{NN})$  dihitung dari  $Tr_j$  untuk setiap kelas dan disusun dalam urutan sesuai dengan ukuran jarak *Euclid*, yaitu:

$$d(\mathbf{x}, \mathbf{y}_{i,j}^{NN}) = \sqrt{(\mathbf{x} - \mathbf{y}_{i,j}^{NN})^T (\mathbf{x} - \mathbf{y}_{i,j}^{NN})} \quad (4)$$

Langkah 2. Hitung vektor rata-rata lokal  $m_{\omega_j}^k$  untuk kelas  $\omega_j$  dengan menggunakan  $k$  tetangga terdekat dalam himpunan  $NN_{\omega_j}^k(\mathbf{x})$

$$m_{\omega_j}^k = \frac{1}{k} \sum_{i=1}^k y_{i,j}^{NN} \quad (5)$$

Langkah 3. Klasifikasikan  $x$  ke dalam kelas  $\omega_c$  ketika jarak *Euclid* antara vektor rata-rata lokal  $m_{\omega_j}^k$  dan  $x$  adalah minimum di antara kelas-kelas  $M$ :

$$\omega_c = \arg \min_{\omega_j} d(\mathbf{x}, m_{\omega_j}^k), j = 1, 2, \dots, M \quad (6)$$

### 2.5 Multi Local Means K-Harmonic Nearest Neighbor (MLM-KHNN)

Pan *et al.* (2016) mengatakan bahwa Metode MLM-KHNN mendeskripsikan klasifikator tetangga terdekat multi lokal yang berbasis  $k$ -harmonik. Tujuan dari aturan

MLM-KHNN adalah untuk meningkatkan kinerja klasifikasi dan mengurangi kepekaan terhadap nilai tunggal dari lingkungan berukuran  $k$  untuk setiap kelas dan nilai seragam  $k$  untuk semua kelas dalam aturan LMKNN. Ide dasar jarak rata-rata harmonik adalah mengambil jarak *Euclid* jumlah rata-rata harmonik antara satu titik data yang diberikan dan setiap titik data dalam kelompok titik yang lain.

Dalam aturan MLM-KHNN label kelas  $\omega_c$  diberikan sampel himpunan  $\mathbf{x} \in R^D$  yang diperoleh sebagai berikut:

Langkah 1: untuk masing-masing kelas  $\omega_j$ , temukan  $k$  tetangga terdekat untuk  $x$  yang dilambangkan dengan  $NN_{\omega_j}^k(x) = \{(y_{i,j}^{NN}, c_{i,j}^{NN})\}_{i=1}^k$  dari  $Tr_j$ , yang kemudian diurutkan dalam urutan sesuai dengan jarak *Euclid* ke  $x$ .

Langkah 2. Hitunglah vektor rata-rata multi lokal  $k$  berdasarkan  $r$  teratas ( $1 \leq r \leq k$ ) tetangga terdekat  $x$  dari  $Tr_j$  di setiap kelas  $\omega_j$ . Untuk kelas  $\omega_j$ , misalkan  $m_{\omega_j}^k(\mathbf{x}) = \{\bar{m}_{\omega_j}^r\}_{r=1}^k$  menunjukkan vektor rata-rata multi-lokal  $k$

$$\bar{m}_{\omega_j}^r = \frac{1}{r} \sum_{i=1}^r y_{i,j}^{NN} \quad (7)$$

Perhatikan bahwa  $\bar{m}_{\omega_j}^r \in R^D$  dan jarak *Euclidnya* yang sesuai ke  $x$  dilambangkan dengan  $d(\mathbf{x}, \bar{m}_{\omega_j}^1), d(\mathbf{x}, \bar{m}_{\omega_j}^2), \dots, d(\mathbf{x}, \bar{m}_{\omega_j}^k)$ .

Langkah 3. Untuk setiap kelas  $\omega_j$ , hitung jarak rata-rata harmonik antara  $x$  dan  $k$  vektor rata-rata multi lokal  $m_{\omega_j}^k(\mathbf{x}) = \{\bar{m}_{\omega_j}^r\}_{r=1}^k$  yang diperoleh dari Langkah 2.

$$HMD(\mathbf{x}, m_{\omega_j}^k(\mathbf{x})) = \frac{k}{\sum_{r=1}^k \frac{1}{d(\mathbf{x}, \bar{m}_{\omega_j}^r)}} \quad (8)$$

Langkah 4. Menetapkan  $x$  ke kelas  $\omega_c$  yang memiliki jarak rata-rata harmonik minimum ke  $x$  dalam Persamaan. (9)

$$\omega_c = \arg \min_{\omega_j} [HMD(\mathbf{x}, m_{\omega_j}^k(\mathbf{x}))] \quad (9)$$

### 3. METODE PENELITIAN

#### 3.1 Data Penelitian

Jenis data yang digunakan dalam penelitian ini merupakan data sekunder. Data sekunder yang dimaksud adalah data penelitian yang diperoleh secara tidak langsung atau melalui perantara (diperoleh dari pihak lain). Data yang digunakan merupakan data bank yang terdaftar di Bank Indonesia serta memiliki kelengkapan data laporan keuangan pada tahun 2017, yang diperoleh dari Bursa Efek Indonesia (BEI).

#### 3.2 Data Training dan Testing

Dilakukan pengujian dengan menggunakan proporsi 90% data latih dan 10% data uji, 80% data latih dan 20% data uji, 70% data latih dan 30% data uji, serta 60% data latih dan 40% data latih. Untuk metode LMKNN, pembagian data latih dan data uji dalam penelitian ini dipilih sebanyak 80% data pertama yakni 69 data sebagai data latih dan 20% data berikutnya sebanyak 18 data sebagai data uji, sedangkan untuk metode MLM-KHNN diambil 80% data pada masing-masing kelas sebagai data latih dan 20% data pada masing-masing kelas sebagai data uji.

#### 3.3 Tahapan Analisis Data

Software yang digunakan dalam penelitian berikut adalah Matlab R2015a dan Microsoft Office Excel 2010. Tahapan dalam analisis data yang diperlukan adalah sebagai berikut:

1. Tahapan analisis data menggunakan metode *Local Mean k-Nearest Neighbor* (LMKNN) sebagai berikut:

- a. Mengelompokkan data ke dalam masing-masing kelas
  - b. Melakukan seleksi variabel dengan menggunakan uji Kruskal Wallis
  - c. Menghitung jarak Mahalanobis sebagai deteksi *outlier*
  - d. Membagi data menjadi 2 bagian, yakni data latih dan data uji, untuk masing-masing kelas
  - e. Menghitung jarak (kedekatan lokasi) antara data latih dan data uji, parameter jarak yang digunakan adalah jarak *euclid*
  - f. Menentukan nilai parameter  $k$
  - g. Mengurutkan nilai jarak tersebut dari nilai terkecil hingga nilai terbesar sesuai  $k$  yang ditetapkan pada masing-masing kelas
  - h. Menentukan jarak terdekat sebanyak nilai  $k$  yang telah ditentukan pada masing-masing kelas
  - i. Menghitung vektor rata-rata lokal  $\mathbf{u}_{ik}$
  - j. Menghitung kembali jarak tiap data uji terhadap nilai vektor rata-rata lokal  $\mathbf{u}_{ik}$
  - k. Memilih nilai jarak terkecil dari setiap kelas yang ada sebagai kelas data uji
  - l. Menghitung APER (*Apparent Error Rate*) hasil klasifikasi
  - m. Melakukan kembali langkah d-k hingga didapati nilai APER (*Apparent Error Rate*) terkecil
2. Tahapan analisis data menggunakan metode *Multi Local Means k-Harmonic Nearest Neighbor* (MLM-KHNN) sebagai berikut:  
Langkah a-h sama seperti LMKNN
- i. Menghitung vektor rata-rata multi lokal  $\mathbf{u}_{ir}$
  - j. Menghitung kembali jarak tiap data uji terhadap nilai vektor rata-rata multi lokal  $\mathbf{u}_{ir}$
  - k. Menghitung nilai jarak rata-rata harmonik
  - l. Memilih nilai jarak rata-rata harmonik terkecil dari setiap kelas yang ada sebagai kelas data uji
  - m. Menghitung APER (*Apparent Error Rate*) hasil klasifikasi
  - n. Melakukan kembali langkah d-k hingga didapati nilai APER (*Apparent Error Rate*) terkecil

Setelah diperoleh hasil klasifikasi dari metode metode LMKNN dan MLM-KHNN, langkah selanjutnya adalah membuat matriks konfusi serta menghitung akurasi klasifikasi dari kedua metode tersebut dengan menggunakan APER untuk menguji metode manakah yang lebih baik digunakan dalam mengklasifikasikan status kesehatan bank di Indonesia.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Seleksi Variabel Menggunakan Uji Kruskal Wallis

Pada bagian ini dilakukan seleksi variabel dengan menggunakan uji Kruskal Wallis. Uji Kruskal Wallis merupakan uji nonparametrik berbasis peringkat yang tujuannya untuk menentukan adakah perbedaan signifikan secara statistik antara dua atau lebih kelompok variabel independen pada variabel dependen yang berskala data numerik (interval atau rasio) dan skala ordinal. Berikut adalah kriteria keputusan hipotesisnya:

Jika  $p\text{-value} < \text{daerah kritis}$  maka keputusan hipotesis adalah menolak  $H_0$  dan menerima  $H_1$  atau yang berarti ada pengaruh variabel bebas terhadap variabel terikat

Jika  $p\text{-value} > \text{daerah kritis}$  maka keputusan hipotesis adalah menerima  $H_0$  dan menolak  $H_1$  atau yang berarti tidak ada pengaruh variabel bebas terhadap variabel terikat

Berikut disajikan nilai  $p\text{-value}$  pada masing-masing variabel yang diperoleh dari *output* program IBM SPSS *Statistics* 20:

**Tabel 1. P-Value Variabel Bebas**

Variabel	P-Value
CAR	0,092
NPL	0,000
LDR	0,000
ROA	0,000
ROE	0,000
NIM	0,000
BOPO	0,000

Berdasarkan Tabel 1, didapatkan  $p\text{-value CAR} = 0,092 > \text{daerah kritis} = 0,05$  yang berarti bahwa tidak ada pengaruh variabel *Capital Adequacy Ratio* (CAR) terhadap status kesehatan bank atau median kedua variabel berbeda secara nyata.  $P\text{-value}$  untuk NPL, LDR, ROA, ROE, NIM, BOPO =  $0,000 < 0,05$  yang berarti bahwa ada pengaruh variabel *Non Performing Loan* (NPL), *Loan to Deposit Ratio* (LDR), *Return on Asset* (ROA), *Return on Equity* (ROE), *Net Interest Margin* (NIM), dan Beban Operasional terhadap Pendapatan Operasional (BOPO) terhadap status kesehatan bank atau median kedua variabel tidak berbeda secara nyata, sehingga variabel yang akan digunakan adalah  $x_1 = \text{NPL}$ ,  $x_2 = \text{LDR}$ ,  $x_3 = \text{ROA}$ ,  $x_4 = \text{ROE}$ ,  $x_5 = \text{NIM}$ , dan  $x_6 = \text{BOPO}$ .

#### 4.2 Jarak Mahalanobis untuk Deteksi *Outlier*

Pada bagian ini diberikan contoh perhitungan manual untuk menghitung jarak Mahalanobis untuk mendeteksi *outlier* multivariat. Disajikan hasil perhitungan dan langkah- langkah dalam menghitung nilai Mahalanobis.

$$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$$

Langkah 1: Mencari vektor  $(\mathbf{x}_j - \bar{\mathbf{x}})^T$

$$(\mathbf{x}_j - \bar{\mathbf{x}}) = \begin{bmatrix} x_{11} - \bar{x}_1 \\ x_{21} - \bar{x}_2 \\ \vdots \\ x_{51} - \bar{x}_5 \\ x_{61} - \bar{x}_6 \end{bmatrix} = \begin{bmatrix} 3,24 - 2,81 \\ 87,74 - 90,83 \\ \vdots \\ 7,68 - 5,35 \\ 80,54 - 86,29 \end{bmatrix}$$

$$(\mathbf{x}_j - \bar{\mathbf{x}}) = \begin{bmatrix} 0,4347 \\ -3,0908 \\ \vdots \\ 2,3314 \\ -5,7517 \end{bmatrix}$$

$$(\mathbf{x}_j - \bar{\mathbf{x}})^T = [0,4347 \quad -3,0908 \quad \dots \quad 2,3314 \quad -5,7517]$$

Langkah 2: Perkalian vektor  $(\mathbf{x}_j - \bar{\mathbf{x}})^T$  dengan invers matriks varian kovarian

$$(\mathbf{x}_j - \bar{\mathbf{x}})^T * \mathbf{S}^{-1} = [0,4347 \quad -3,0908 \quad \dots \quad 2,3314 \quad -5,7517] * \begin{bmatrix} 0,4101 & 0,0026 & \dots & -0,1576 & 0,0125 \\ 0,0026 & \ddots & & & 0,0017 \\ \vdots & & \ddots & & \vdots \\ -0,1576 & & & \ddots & -0,0401 \\ -0,0125 & 0,0017 & \dots & -0,0410 & 0,0286 \end{bmatrix} = [0,3271 \quad 0,0003 \quad \dots \quad 0,2241 \quad 0,1550]$$

Langkah 3: Menghitung jarak Mahalanobis

$$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$$

$$= [0,3271 \quad 0,0003 \quad \dots \quad 0,2241 \quad 0,1550] * \begin{bmatrix} 0,4347 \\ -3,0908 \\ \vdots \\ 2,3314 \\ -5,7517 \end{bmatrix}$$

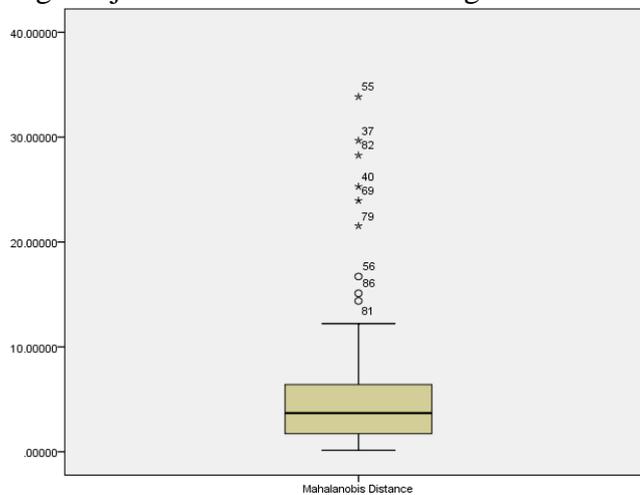
$$d_1^2 = 4,7043$$

$$d_2^2 = 0,8717$$

⋮

$$d_{87}^2 = 6,1160$$

Cara mengidentifikasi terjadinya *outlier* multivariat salah satunya adalah dengan menggunakan *box plot*. Berikut adalah *box plot* yang didapatkan dari hasil perhitungan jarak Mahalanobis yang disajikan dalam Gambar 1 sebagai berikut:



**Gambar 4.** *Box Plot* Jarak Mahalanobis

Berdasarkan Gambar 1, didapatkan kesimpulan bahwa terdapat 9 *outlier* atau pencilan pada data tersebut. *Outlier* terdapat pada data ke 37, 40, 55, 56, 69, 79, 81, 82 dan 86.

#### 4.3 Metode *Local Mean K-Nearest Neighbor* (LMKNN)

Disajikan nilai APER yang telah diujicobakan ke semua proporsi data latih dan data uji dari program Matlab R2015a, sehingga diperoleh nilai APER dalam sebagai berikut:

**Tabel 2.** Nilai APER untuk Setiap Proporsi Metode LMKNN

K	Proporsi Data Latih : Data Uji			
	90:10	80:20	70:30	60:40
1	0,6667	0,6667	0,6538	0,6286
3	0,2222	0,1111	0,2308	0,2857
5	0,3333	0,1667	0,1923	0,2571
7	0,1111	0,0556	0,1538	-
9	0,1111	-	-	-

Keterangan: - = tidak dapat dicobakan karena nilai  $k >$  banyaknya data latih ( $n$ )

Berdasarkan Tabel 2, diperoleh nilai APER terkecil, yaitu 0,0556 dengan proporsi data latih dan data uji 80:20 dengan nilai  $k = 7$ , yang berarti bahwa nilai APER terkecil terdapat pada proporsi data latih sebanyak 80% dan data uji sebanyak 20% dengan jumlah tetangga terdekat atau  $k = 7$ , sehingga terpilih dengan menggunakan nilai  $k = 7$ .

Jarak *Euclid* =  $(\mathbf{x}_i, \mathbf{y}_j) = \sqrt{\sum_{l=1}^p (x_{lp} - y_{jp})^2}$ ; dengan  $\mathbf{x}_i$  adalah data uji, dan  $\mathbf{y}_j$  adalah data latih.

$$d_{18,1} = \sqrt{(5,62 - 3,24)^2 + (94,55 - 87,74)^2 + \dots + (123,00 - 80,54)^2} \\ = 46,3224$$

$$d_{18,2} = \sqrt{(5,62 - 1,70)^2 + (94,55 - 82,10)^2 + \dots + (123,00 - 82,86)^2} \\ = 44,6383$$

⋮

$$d_{18,69} = \sqrt{(5,62 - 1,98)^2 + (94,55 - 124,01)^2 + \dots + (123,00 - 128,87)^2} \\ = 31,3239$$

Hasil perhitungan jarak *Euclid* data uji ke semua data latih yang diurutkan dari nilai terkecil hingga nilai terbesar. Kemudian dipilih 7 amatan dengan jarak terdekat pada masing-masing kelas untuk menghitung nilai vektor rata-rata lokal.

Vektor rata-rata lokal kelas 1 (sangat sehat):  $\mathbf{u}_1 = \begin{bmatrix} 1,89 \\ 90,41 \\ 1,99 \\ 12,23 \\ 5,90 \\ 81,71 \end{bmatrix}$ ; Vektor rata-rata lokal kelas 2

(sehat):  $\mathbf{u}_2 = \begin{bmatrix} 2,62 \\ 90,95 \\ 1,52 \\ 7,02 \\ 4,78 \\ 87,00 \end{bmatrix}$ ; Vektor rata-rata lokal kelas 3 (cukup sehat):  $\mathbf{u}_3 = \begin{bmatrix} 3,31 \\ 92,12 \\ 0,53 \\ 2,80 \\ 5,10 \\ 95,56 \end{bmatrix}$ ; Vektor

rata-rata lokal kelas 4 (kurang sehat):  $\mathbf{u}_4 = \begin{bmatrix} 6,36 \\ 92,01 \\ -0,71 \\ -5,84 \\ 5,11 \\ 109,57 \end{bmatrix}$

Setelah diperoleh nilai vektor rata-rata lokal pada masing-masing kelas, selanjutnya dilakukan perhitungan jarak *Euclid* untuk prediksi kelas dari data uji sebagai berikut:

Jarak *Euclid* =  $d(\mathbf{x}, \mathbf{u}) = \sqrt{\sum_{i=1}^p (x_i - u_i)^2}$  dengan  $\mathbf{x}$  adalah data uji dan  $\mathbf{u}$  adalah vektor rata-rata lokal.

a. Jarak *Euclid* pada kelas 1 (sangat sehat)

$$d(\mathbf{x}, \mathbf{u}_1) = 46,1112$$

b. Jarak *Euclid* pada kelas 2 (sehat)

$$d(\mathbf{x}, \mathbf{u}_2) = 39,0990$$

c. Jarak *Euclid* pada kelas 3 (cukup sehat)

$$d(\mathbf{x}, \mathbf{u}_3) = 29,4717$$

d. Jarak *Euclid* pada kelas 4 (kurang sehat)

$$d(\mathbf{x}, \mathbf{u}_4) = 13,8384$$

Dari keempat nilai jarak *Euclid* pada masing-masing kelas, data uji diprediksi masuk ke dalam kelas dengan nilai *Euclid* terkecil, yaitu 13,8384 dengan label kelas 4 (kurang sehat).

Berikut adalah tabel matriks konfusi untuk metode LMKNN:

**Tabel 3.** Matriks Konfusi LMKNN

		Prediksi				Ketepatan Klasifikasi (%)
		SS	S	CS	KS	
Aktual	SS	8	0	0	0	100
	S	0	3	1	0	75
	CS	0	0	4	0	100
	KS	0	0	0	2	100

#### 4.4 Metode *Multi Local Means K-Harmonic Nearest Neighbor* (MLM-KHNN)

Disajikan nilai *Apparent Error Rate* (APER) terkecil yang diperoleh dari *output* program Matlab R2015a yang telah diujicobakan ke semua proporsi data latih dan data uji, sehingga diperoleh nilai APER dalam Tabel 4 sebagai berikut:

**Tabel 4.** Nilai APER untuk Setiap Proporsi Metode MLM-KHNN

<i>k</i>	Proporsi Data Latih : Data Uji			
	90:10	80:20	70:30	60:40
1	0,3333	0,2222	0,2692	0,2571
3	0,2222	0,1667	0,2308	0,3143
5	0,3333	0,1667	0,2308	0,2857
7	0,3333	0,1667	0,2308	-
9	0,3333	-	-	-

Keterangan: - = tidak dapat dicobakan karena nilai *k* > banyaknya data latih (*n*)

Berdasarkan Tabel 25, diperoleh nilai APER terkecil, yaitu 0,1667 dengan proporsi data latih dan data uji 80:20 dengan nilai *k* = 3, yang berarti bahwa nilai APER terkecil terdapat pada proporsi data latih sebanyak 80% dan data uji sebanyak 20% dengan jumlah tetangga terdekat atau *k* = 3, sehingga dilakukan contoh perhitungan manual dengan menggunakan nilai *k* = 3.

Jarak *Euclid* =  $(\mathbf{x}_i, \mathbf{y}_j) = \sqrt{\sum_{l=1}^p (x_{lp} - y_{jp})^2}$ ; dengan  $\mathbf{x}_i$  adalah data uji, dan  $\mathbf{y}_j$  adalah data latih.

$$d_{18,1} = 46,3224$$

$$d_{18,2} = 44,6383$$

⋮

$$d_{18,69} = 31,3239$$

Hasil perhitungan jarak *Euclid* data uji ke semua data latih yang diurutkan dari nilai terkecil hingga nilai terbesar. Kemudian dipilih 3 amatan dengan jarak terdekat pada masing-masing kelas untuk menghitung nilai vektor rata-rata multi lokal.

Vektor rata-rata multi lokal kelas 1 (sangat sehat)  $\mathbf{u}_1$ :

	r = 1	r = 2	r = 3
$y_1$	1,52	1,61	2,08
$y_2$	97,14	89,62	92,83
$y_3$	1,60	1,67	1,73
$y_4$	6,30	6,63	9,05
$y_5$	4,95	4,88	5,43
$y_6$	83,34	83,10	83,20

Vektor rata-rata multi lokal kelas 2 (sehat)  $\mathbf{u}_2$ :

	r = 1	r = 2	r = 3
$y_1$	0,00	1,30	2,13
$y_2$	114,53	101,43	94,48
$y_3$	2,05	1,75	1,59
$y_4$	4,97	5,31	6,04
$y_5$	3,11	3,44	4,44
$y_6$	92,52	89,50	89,31

Vektor rata-rata multi lokal kelas 3 (cukup sehat)  $\mathbf{u}_3$ :

	r = 1	r = 2	r = 3
$y_1$	1,09	3,60	3,59
$y_2$	83,57	83,23	85,21
$y_3$	0,32	0,32	0,39
$y_4$	0,70	1,21	2,31
$y_5$	3,85	4,50	4,94
$y_6$	97,81	97,18	96,69

Vektor rata-rata multi lokal kelas 4 (kurang sehat)  $\mathbf{u}_4$ :

	r = 1	r = 2	r = 3
$y_1$	5,37	5,97	6,75
$y_2$	91,95	92,97	86,21
$y_3$	-1,43	-1,17	-1,27
$y_4$	-15,43	-10,35	-8,99
$y_5$	3,07	4,60	4,67
$y_6$	117,66	113,04	113,26

Setelah diperoleh nilai vektor rata-rata multi lokal pada masing-masing kelas, selanjutnya yaitu mencari mencari jarak rata-rata harmonik. Langkah-langkahnya yaitu terlebih dahulu dilakukan perhitungan jarak *Euclid* yang baru. Jarak *Euclid* =  $d(\mathbf{x}, \mathbf{u}) =$

$$\sqrt{\sum_{i=1}^p (x_i - u_i)^2}$$

dengan  $\mathbf{x}$  adalah data uji dan  $\mathbf{u}$  adalah vektor rata-rata multi lokal.

a. Jarak *Euclid* pada kelas 1 (sangat sehat)

$$\text{untuk } r=1: d(\mathbf{x}, \mathbf{u}_{1_1}) = 42,2915; r=2: d(\mathbf{x}, \mathbf{u}_{1_2}) = 42,8234; r=3: d(\mathbf{x}, \mathbf{u}_{1_3}) = 43,2980$$

b. Jarak *Euclid* pada kelas 2 (sehat)

$$\text{untuk } r=1: d(\mathbf{x}, \mathbf{u}_{2_1}) = 39,0220; r=2: d(\mathbf{x}, \mathbf{u}_{2_2}) = 36,8276; r=3: d(\mathbf{x}, \mathbf{u}_{2_3}) = 36,4931$$

c. Jarak *Euclid* pada kelas 3 (cukup sehat)

$$\text{untuk } r=1: d(\mathbf{x}, \mathbf{u}_{3_1}) = 29,0103; r=2: d(\mathbf{x}, \mathbf{u}_{3_2}) = 29,5430; r=3: d(\mathbf{x}, \mathbf{u}_{3_3}) = 29,6253$$

d. Jarak *Euclid* pada kelas 4 (kurang sehat)

$$\text{untuk } r=1: d(\mathbf{x}, \mathbf{u}_{4_1}) = 10,3847; r=2: d(\mathbf{x}, \mathbf{u}_{4_2}) = 10,6724; r=3: d(\mathbf{x}, \mathbf{u}_{4_3}) = 13,0582$$

Langkah terakhir yaitu melakukan prediksi kelas dari data uji menggunakan jarak rata-

rata harmonik pada masing-masing kelas, dengan rumus:  $\text{HMD}(\mathbf{x}, \mathbf{u}_{\omega_j}^k(x)) = \frac{k}{\sum_{r=1}^k \frac{1}{d(x, u_{\omega_j}^r)}}$

a. Jarak rata-rata harmonik kelas 1

$$\text{HMD}_1 = \frac{5}{\left(\frac{1}{42,2915}\right) + \left(\frac{1}{42,8234}\right) + \left(\frac{1}{43,2980}\right)} = 42,8004$$

b. Jarak rata-rata harmonik kelas 2

$$\text{HMD}_2 = \frac{5}{\left(\frac{1}{39,0220}\right) + \left(\frac{1}{36,8276}\right) + \left(\frac{1}{36,4931}\right)} = 37,4146$$

c. Jarak rata-rata harmonik kelas 3

$$HMD_3 = \frac{5}{\left(\frac{1}{29,0103}\right) + \left(\frac{1}{29,5430}\right) + \left(\frac{1}{29,6253}\right)} = 29,3903$$

d. Jarak rata-rata harmonik kelas 4

$$HMD_4 = \frac{5}{\left(\frac{1}{10,3847}\right) + \left(\frac{1}{10,6724}\right) + \left(\frac{1}{13,0582}\right)} = 11,2538$$

Dari keempat nilai jarak rata-rata harmonik pada masing-masing kelas, data uji diprediksi masuk ke dalam kelas dengan nilai rata-rata harmonik terkecil, yaitu 11,2538 dengan label kelas 4 (kurang sehat).

Berikut adalah tabel matriks konfusi untuk metode MLM-KHNN:

**Tabel 5.** Matriks Konfusi MLM-KHNN

		Prediksi				Ketepatan Klasifikasi (%)
		SS	S	CS	KS	
Aktual	SS	7	1	0	0	87,5
	S	1	2	1	0	50
	CS	0	0	4	0	100
	KS	0	0	0	2	100

## 5. PENUTUP

### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dipaparkan pada bab sebelumnya, maka dapat diambil kesimpulan sebagai berikut:

1. Setelah melakukan seleksi variabel, dari tujuh variabel bebas terdapat satu variabel yang tidak memiliki pengaruh yang signifikan terhadap status kesehatan bank sebagai variabel independen atau variabel tak bebas. Tujuh variabel tersebut adalah variabel *Capital Adequacy Ratio* (CAR), *Non Performing Loan* (NPL), *Loan to Deposit Ratio* (LDR), *Return on Asset* (ROA), *Return on Equity* (ROE), *Net Interest Margin* (NIM), dan Beban Operasional terhadap Pendapatan Operasional (BOPO). Variabel bebas yang tidak memiliki pengaruh yaitu variabel CAR, sedangkan variabel NPL, LDR, ROA, ROE, NIM, dan BOPO memiliki pengaruh yang signifikan terhadap status kesehatan bank.
2. Pada deteksi *outlier* menggunakan jarak Mahalanobis, terdapat 9 *outlier* atau pencilan pada data kesehatan bank yang digunakan. *Outlier* terdapat pada data ke 37, 40, 55, 56, 69, 79, 81, 82 dan data ke 86.
3. Nilai akurasi menggunakan program bantuan Matlab2015a pada metode LMKNN menghasilkan nilai  $k$  optimal pada proporsi data latih sebesar 80% dan data data uji sebesar 20% dengan jumlah tetangga terdekat atau  $k = 7$  dengan nilai *Apparent Error Rate* (APER) sebesar 0,0556 dan nilai akurasi sebesar 94,44%. Pada metode MLM-KHNN, menghasilkan nilai  $k$  optimal pada proporsi data latih sebesar 80% dan data data uji sebesar 20% dengan jumlah tetangga terdekat atau  $k = 3$  dengan nilai APER sebesar 0,1667 dan nilai akurasi sebesar 83,33%.
4. Berdasarkan perhitungan nilai APER dapat disimpulkan bahwa klasifikasi menggunakan metode LMKNN lebih baik dibandingkan metode MLM-KHNN dalam mengklasifikasikan status kesehatan bank di Indonesia, sehingga terbukti bahwa metode LMKNN dapat melakukan klasifikasi lebih akurat disaat terdapat *outlier* pada data penelitian yang digunakan.

### 5.2 Saran

Untuk penelitian selanjutnya, pemilihan proporsi data sebaiknya menggunakan metode *k-fold* agar menghasilkan proporsi data serta klasifikasi yang lebih akurat dan valid.

#### DAFTAR PUSTAKA

- [BI] Bank Indonesia. 1998. *Undang-Undang Nomor 10 Tahun 1998 tentang Perbankan*. Jakarta: Bank Indonesia.
- [BI] Bank Indonesia. 2004. *Surat Edaran Bank Indonesia No.6/23/DPNP Tahun 2004 tentang Sistem Penilaian Tingkat Kesehatan Bank Umum*. Jakarta: Direktorat Hukum.
- [BI] Bank Indonesia. 2011. *Surat Edaran Bank Indonesia No.13/24/DPNP Tahun 2011 tentang Penilaian Tingkat Kesehatan Bank Umum*. Jakarta: Direktorat Penelitian dan Pengaturan Perbankan.
- Bezdek, J.C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press.
- Dendawijaya, L. 2005. *Manajemen Perbankan*. Edisi Kedua. Bogor: Ghalia Indonesia.
- Gorunescu, F. 2011. *Data Mining: Concepts, Model and Techniques*. Jerman: Springer.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. 2010. *Multivariate Data Analysis*. 7<sup>th</sup> ed. New Jersey: Prentice Hall, Inc.
- Han, J., Kamber, M. 2006. *Data Mining: Concepts and Techniques*. 2<sup>nd</sup> ed. San Fransisco: Morgan Kaufmann Publisher.
- Han, J., Kamber, M. dan Pei, J. 2012. *Data Mining: Concepts and Techniques*. 3<sup>rd</sup> ed. Waltham: Morgan Kaufmann Publishers.
- Hand, D.J., Mannila, H., Smyth, P. 2001. *Principles of Data Mining*. Cambridge: MIT Press.
- [IAI] Ikatan Akuntansi Indonesia. 2007. *Standar Akuntansi Keuangan*. Jakarta: Salemba Empat.
- Johnson, R.A., Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Prentice Hall.
- Kaligis, Y.W. 2013. *Analisis Tingkat Kesehatan Bank dengan Menggunakan Metode CAMEL pada Industri Perbankan BUMN yang Terdaftar di Bursa Efek Indonesia*. Jurnal EMBA Vol. 1, No. 3: Hal. 263-272.
- Kasmir. 2008. *Analisis Laporan Keuangan*. Jakarta: Rajawali Pers.
- Koch, T.W., MacDonald, S.S. 2000. *Bank Management*. 4<sup>th</sup> ed. Orlando: Dryden Press.
- Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Mitani, Y., Hamamoto, Y. 2006. *A Local Mean-Based Nonparametric Classifier*. Pattern Recognition Letters Vol. 27, No. 10: Hal. 1151-1159.
- Pan, Z., Wang, Y. dan Ku W. 2016. *A New k-Harmonic Nearest Neighbor Classifier Based on the Multi-Local Means*. Pattern Recognition Letters No.67: Hal. 115-125.
- Prasetyo, E. 2012. *Data Mining-Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.
- Prasetyo, E. 2014. *Data Mining-Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi.
- Priyatno. 2013. *Analisis data dengan SPSS*. Jakarta: Media Kom.
- Rose, P.S. 2002. *Commercial Bank Management*. Boston: McGraw-Hill Irwin.
- Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Syaifuddin, D.T. 2007. *Manajemen Perbankan*. Kendari: Unhalu Press.

Tan, P., Steinbach, M., Kumar, V. 2006. *Introduction to Data Mining*. Boston: Pearson Education.