

## KLASIFIKASI CALON DEBITUR KREDIT PEMILIKAN RUMAH (KPR) MULTIGUNA TAKE OVER MENGGUNAKAN METODE $k$ NEAREST NEIGHBOR DENGAN PEMBOBOTAN GLOBAL GINI DIVERSITY INDEX

Inas Hasimah<sup>1</sup>, Moch. Abdul Mukid<sup>2</sup>, Hasbi Yasin<sup>3</sup>

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

[mamukid@yahoo.com](mailto:mamukid@yahoo.com)

### ABSTRACT

House credit (KPR) is a credit facilities for buying or other comsumptive needs with house warranty. The warranty for KPR is the house that will be purchased. The warranty for KPR multiguna take over is the house that will be owned by debtor, and then debtor is taking over KPR to another financial institution. For fulfilled the credit to prospective debtor is done by passing through the process of credit application and credit analysis. With the credit analysis, will acknowledge the ability of debtor for repay a credit. Final decision of credit application is classified into approved and refused.  $k$  Nearest Neighbor by attributes weighting using Global Gini Diversity Index is a statistical method that can be used to classify the credit decision of prospective debtor. This research use 2443 data of KPR multiguna take over's prospective debtor in 2018 with credit decision of prospective debtor as dependent variable and four selected independent variable such as home ownership status, job, loans amount, and income. The best classification result of  $k$ -NN by Global Gini Diversity Index weighting is when using 80% training data set and 20% testing data set with  $k=7$  obtained APER value 0,0798 and accuracy 92,02%.

**Keywords:** KPR Multiguna Take Over, Classification, KNN by Global Gini Diversity Index weighting, Evaluation of Classification

### 1. PENDAHULUAN

Rumah merupakan kebutuhan dasar setiap manusia. Ketersediaan lahan semakin berkurang seiring dengan bertambahnya populasi penduduk dan mengakibatkan harga beli tanah maupun rumah menjadi tinggi. Menurut Ayodya (2016) keadaan tersebut menyebabkan masyarakat yang berpenghasilan rendah mempunyai kemungkinan kecil untuk membeli rumah secara tunai. Sehingga salah satu alternatif yang dapat ditempuh oleh pemerintah dalam rangka mewujudkan masyarakat yang adil dan makmur adalah dengan kebijakan pemberian kredit untuk kepemilikan rumah dan menunjuk bank maupun lembaga keuangan non bank sebagai penyalur Kredit Pemilikan Rumah (KPR).

Kredit Pemilikan Rumah (KPR) adalah kredit yang digunakan untuk membeli atau untuk kebutuhan konsumtif lainnya dengan jaminan atau agunan berupa rumah. Agunan yang diperlukan untuk KPR adalah rumah yang akan dibeli itu sendiri untuk KPR pembelian. Sedangkan dalam KPR Multiguna atau KPR *Refinancing* yang menjadi agunan adalah rumah yang sudah dimiliki (Kuswati, 2015). Selama jangka waktu berjalannya KPR, terkadang adakalanya nasabah melakukan permohonan KPR di bank lain atau lembaga keuangan lain, kondisi tersebut disebut dengan peralihan kredit (*take over*). Dalam kaitannya dengan KPR Multiguna *take over*, peralihan atas pemindahan kewajiban yang dimaksud adalah pembiayaan untuk pengambilalihan kredit dari KPR Multiguna

bank lain yang telah berjalan selama jangka waktu tertentu. Walaupun sebagian besar nasabah menyatakan cukup puas terhadap tingkat suku bunga yang dibebankan kepada mereka, namun 60% nasabah menyatakan bahwa besarnya tingkat suku bunga tetap menjadi perhatian utama nasabah ketika akan mengajukan pinjaman (Direktorat Kredit, BPR, dan UMKM BI, 2009). Oleh karena itu, pengajuan untuk KPR Multiguna *take over* dapat disebabkan karena nasabah ingin memperoleh tingkat suku bunga yang lebih rendah.

Penelitian ini bertujuan untuk mengetahui atribut-atribut yang berhubungan dengan keputusan akhir kredit serta mengklasifikasikan data calon debitur KPR Multiguna *take over* tahun 2018 pada Bank X di Provinsi DKI Jakarta menggunakan metode klasifikasi *k Nearest Neighbor* dengan pembobotan *Global Gini Diversity Index*.

## **2. TINJAUAN PUSTAKA**

### **2.1. Kredit**

Menurut Pasal 1 angka 11 Undang-Undang Republik Indonesia Nomor 10 Tahun 1998, kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga.

Pemberian suatu fasilitas kredit mempunyai beberapa tujuan yang hendak dicapai yang tentunya tergantung dari tujuan bank itu sendiri. Tujuan pemberian kredit juga tidak akan terlepas dari misi bank tersebut didirikan. Dalam praktiknya tujuan pemberian suatu kredit sebagai berikut (Kasmir, 2010) :

1. Mencari keuntungan
2. Membantu usaha nasabah
3. Membantu pemerintah

### **2.2. Kredit Pemilikan Rumah (KPR)**

Cara pendanaan yang paling umum ditempuh para calon pembeli rumah adalah dengan fasilitas Kredit Pemilikan Rumah (KPR) atau hipotik konvensional. Hipotik adalah suatu bentuk ikatan perjanjian antara pembeli rumah (peminjam uang) dan yang memberi pinjaman uang, biasanya bank. (Tim Penulis Indira, 1996).

### **2.3. Kredit Pemilikan Rumah (KPR) Multiguna *Take Over***

Kredit Pemilikan Rumah (KPR) dan Kredit Multiguna termasuk dalam kredit konsumsi. Kredit konsumsi adalah pemberian kredit untuk keperluan konsumsi, keperluan konsumsi yang dimaksud adalah pemenuhan akan kebutuhan hidup termasuk kebutuhan untuk tempat tinggal, dalam hal ini adalah rumah. Menurut Kementerian Pekerjaan Umum dan Perumahan Rakyat (2017), Pembiayaan multiguna adalah pembiayaan untuk pengadaan barang dan/atau jasa yang diperlukan oleh debitur untuk pemakaian/konsumsi dan bukan untuk keperluan usaha (aktivitas produktif) dalam jangka waktu yang diperjanjikan. Sehingga Kredit Pemilikan Rumah (KPR) Multiguna adalah kredit perorangan yang diajukan untuk memenuhi kebutuhan hidup. Salah satu bentuk dari *refinancing* adalah peralihan kredit atau biasa dikenal dengan *take over* kredit. *Take over* kredit adalah pemindahan atas kewajiban dari kreditur satu kepada kreditur lain. Dalam kaitannya dengan KPR Multiguna *Take Over*, pemindahan kewajiban yang dimaksud adalah pembiayaan untuk pengambilalihan kredit dari KPR Multiguna bank lain yang telah berjalan selama jangka waktu tertentu.

## 2.4. Data Mining

Menurut Gorunescu (2011), secara sistematis ada tiga langkah utama dalam *data mining*, yaitu :

1. Eksplorasi / pemrosesan awal data
2. Membangun model dan melakukan validasi terhadapnya
3. Penerapan

## 2.5. Seleksi Atribut menggunakan Uji Independensi

Menurut Han dan Kamber (2006), dalam *data mining* seringkali membutuhkan *data integration*, yaitu penggabungan data dari banyak tempat-tempat penyimpanan data. Dalam *data integration*, redundansi menjadi salah satu masalah penting. Beberapa redundansi dapat dideteksi dengan uji independensi. Dicontohkan dua atribut, dengan menggunakan uji independensi dapat mengukur kekuatan satu atribut menyiratkan atribut yang lain berdasarkan data yang ada. Untuk data kategorik (diskrit), suatu hubungan korelasi antara dua atribut, A dan B, dapat dicari dengan uji independensi menggunakan statistik uji  $\chi^2$  (chi-square). Misalkan A memiliki nilai yang berbeda sebanyak c, dinamakan  $a_1, a_2, \dots, a_c$ . B memiliki nilai yang berbeda sebanyak r, dinamakan  $b_1, b_2, \dots, b_r$ . Tupel data yang dideskripsikan oleh A dan B dapat ditunjukkan dalam tabel kontingensi, dengan nilai sebanyak c dari A membentuk kolom-kolom dan nilai sebanyak r dari B membentuk baris-baris. Tampilan tabel kontingensi yang dimaksud adalah sebagai berikut :

**Tabel 1.** Tabel Kontingensi  $r \times c$

	$a_1$	$a_{\dots}$	$a_c$	Total
$b_1$	$n_{11}$	$n_{\dots 1}$	$n_{c1}$	$n_{11} + n_{\dots 1} + n_{c1}$
$b_2$	$n_{12}$	$n_{\dots 2}$	$n_{c2}$	$n_{12} + n_{\dots 2} + n_{c2}$
$b_{\dots}$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$b_r$	$n_{1r}$	$n_{\dots r}$	$n_{cr}$	$n_{1r} + n_{\dots r} + n_{cr}$
Total	$n_{11} + n_{12} + \dots + n_{1r}$	$n_{\dots 1} + n_{\dots 2} + \dots + n_{\dots r}$	$n_{c1} + n_{c2} + \dots + n_{cr}$	$N$

Misalkan  $(A_i, B_j)$  menyatakan kejadian bahwa atribut A memiliki nilai  $a_i$  dan atribut B nilai  $b_j$  atau dapat dituliskan di mana  $(A = a_i, B = b_j)$ . Tiap dan setiap kemungkinan kejadian bersama  $(A_i, B_j)$  akan memiliki sel tersendiri pada tabel kontingensi. Nilai  $\chi^2$  (juga dikenal dengan Pearson  $\chi^2$  statistic) dihitung dengan rumus berikut :

$$\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \quad (1)$$

dengan  $o_{ij}$  adalah frekuensi observasi (perhitungan aktual) dari kejadian bersama  $(A_i, B_j)$  dan  $e_{ij}$  adalah frekuensi harapan dari  $(A_i, B_j)$ ,  $e_{ij}$  dapat dihitung dengan :

$$e_{ij} = \frac{\text{count}(A = a_i) \times \text{count}(B = b_j)}{N} \quad (2)$$

di mana N adalah jumlah tupel data,  $\text{count}(A = a_i)$  adalah jumlah tupel yang memiliki nilai  $a_i$  untuk A, dan  $\text{count}(B = b_j)$  adalah jumlah tupel yang memiliki nilai  $b_j$  untuk B.

Jumlahan pada persamaan (1) dihitung dari nilai-nilai yang terletak pada sel  $r \times c$ . Perlu diperhatikan bahwa sel-sel yang paling berkontribusi untuk nilai  $\chi^2$  adalah jika nilai aktualnya berbeda dengan nilai harapannya. Uji independensi menggunakan statistik uji  $\chi^2$  memiliki hipotesis bahwa atribut A dan atribut B adalah independen. Uji ini didasarkan pada suatu tingkat signifikansi, dengan derajat kebebasan  $(r - 1) \times (c - 1)$ . Jika hipotesis ditolak, maka dapat dikatakan bahwa atribut A dan atribut B secara statistik berhubungan.

## 2.6. Klasifikasi

Menurut Prasetyo (2012), klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototipe untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan / klasifikasi / prediksi pada suatu objek data lain agar diketahui kelas dari objek data tersebut dalam model yang sudah disimpannya.

## 2.7. *k*-Nearest Neighbor (k-NN)

Menurut Prasetyo (2012), *Nearest Neighbor* merupakan teknik klasifikasi yang sederhana, tetapi mempunyai hasil kerja yang cukup bagus. Meskipun begitu, k-NN juga mempunyai kelebihan dan kekurangan. Beberapa karakteristik k-NN adalah sebagai berikut :

1. Karena k-NN masuk kategori *lazy learning* yang menyimpan sebagian atau semua data dan hampir tidak ada proses pelatihan, k-NN sangat cepat dalam proses pelatihan (karena memang tidak ada), tetapi sangat lambat dalam proses prediksi.
2. Hal yang rumit adalah menentukan nilai k yang paling sesuai.
3. Karena k-NN pada prinsipnya memilih tetangga terdekat, parameter jarak juga penting untuk dipertimbangkan sesuai dengan kasus datanya. Euclidean sangat cocok untuk menggunakan jarak terdekat (lurus) antara dua data, tetapi Manhattan sangat teguh (*robust*) untuk mendeteksi *outlier* dalam data.

## 2.8. *Attributes Weighting in Nearest Neighbor Classification*

Menurut Chen dan Guo (2015), dalam penerapan k-NN untuk melakukan klasifikasi terhadap data kategorik, diperlukan cara khusus dalam menghitung jarak antar data yang ada. Salah satu cara khusus dalam menghitung jarak antar data yang paling sering diterapkan adalah menggunakan *Simple Matching Coefficient* (SMC).

Dalam penelitian ini digunakan sebuah metode klasifikasi terhadap data kategorik yang menggunakan metode klasifikasi *nearest neighbor* dengan pembobotan atribut menggunakan jarak *Weighted Simple Matching Coefficient* (WSMC).

### 2.8.1. *Simple Matching Coefficient* (SMC)

Misalkan terdapat kumpulan data *training*  $tr = \{z_1, z_2, \dots, z_N\}$  dengan  $z_i = (x_i, y_i)$  menjadi sampel ke- $i$  dengan  $i = 1, 2, \dots, N$  yang mana  $N$  menyatakan banyaknya sampel. Pada bagian ini  $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$  adalah objek data yang dilengkapi dengan  $D$  sebagai atribut kategorik dan  $y_i$  kelas asli dari  $x_i$  dengan  $y_i \in \{1, 2, \dots, M\}$  dengan  $M$  adalah jumlah kelas yang mengandung kumpulan data *training*. Variabel  $m$  menunjukkan kelas, dengan  $m = 1, 2, \dots, M$ . Kumpulan untuk kategori-kategori yang ada dilambangkan dengan  $S_d$ , sementara suatu kategori yang terdapat pada  $S_d$  dilambangkan dengan  $s_d \in S_d$  (Chen dan Guo, 2015).

Penghitungan jarak menggunakan *Simple Matching Coefficient* (SMC) yang didefinisikan sebagai berikut :

$$SMC(\mathbf{x}_i, \mathbf{x}_j) = \sum_{d=1}^D I(x_{id} \neq x_{jd}) \quad (3)$$

dengan  $I(\cdot)$  adalah fungsi indikator dengan  $I(\text{true}) = 1$  dan  $I(\text{false}) = 0$ .

### 2.8.2 *Weighted Simple Matching Coefficient* (WSMC) dengan Metode pembobotan *Global Gini Diversity Index*

*Weighted Simple Matching Coefficient* (WSMC) adalah perhitungan statistik yang hampir sama dengan *Simple Matching Coefficient* yang melakukan perbandingan untuk tiap atribut dengan atribut yang lain, namun dalam implementasinya setiap atribut tidak akan dihitung secara rata, namun dengan memberikan bobot terhadap tiap atribut sehingga secara tidak langsung mengurangi pengaruh data yang akan mengacaukan klasifikasi. Metode penghitungan ini dapat digunakan secara efektif untuk melakukan klasifikasi terhadap data yang bersifat kategorik (Chen dan Guo, 2015).

Pengukuran jarak baru akan dinotasikan dengan  $WSMC_{global}$  sesuai dengan metode pembobotan global yang digunakan. Pada metode global, atribut dikaitkan dengan vektor bobot  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_d, \dots, \omega_D)$  dan jarak WSMC antara objek data  $\mathbf{x}_i$  dan  $\mathbf{x}_j$  diberikan pada persamaan (Chen dan Guo, 2015) :

$$WSMC_{global}(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\omega}) = \sum_{d=1}^D \omega_d^{(GG)} \times I(x_{id} \neq x_{jd}) \quad (4)$$

Bobot  $\omega_d$  didefinisikan dengan :

$$\omega_d^{(GG)} = \exp \left[ -\frac{M}{M-1} \left( \sum_{s_d \in S_d} p(s_d) \times GG(s_d) \right) \right] \quad (5)$$

Indeks global gini rata-rata dari atribut ke- $d$  dirumuskan dengan :

$$AGG(d) = \frac{M}{M-1} \left( \sum_{s_d \in S_d} p(s_d) \times GG(s_d) \right) \quad (6)$$

Sehingga persamaan (5) dapat ditulis menjadi sebagai berikut :

$$\omega_d^{(GG)} = \exp[-(AGG(d))] \quad (7)$$

$GG(s_d)$  dihitung dengan :

$$GG(s_d) = 1 - \sum_{m=1}^M [p(m|s_d)]^2 \quad (8)$$

$p(m|s_d)$  adalah peluang untuk perhitungan derajat kontribusi tiap atribut terhadap tiap kelas.  $p(m|s_d)$  dihitung dengan :

$$p(m|s_d) = \frac{\sum_{(x,y) \in c_m} I(x_d = s_d)}{\sum_{(x,y) \in tr} I(x_d = s_d)} \quad (9)$$

Selanjutnya berikut adalah persamaan untuk menghitung  $p(s_d)$  adalah :

$$p(s_d) = \frac{1}{N} \sum_{(x,y) \in tr} I(x_d = s_d) \quad (10)$$

$N$  menunjukkan banyaknya data. Kelas untuk data *testing* yang terpilih adalah berdasarkan :

$$y = \operatorname{argmax}_m \sum_{(x_j, y_j) \in NN_z} I(m = y_j) \quad (11)$$

## 2.9. Evaluasi Ketepatan Hasil Klasifikasi

Menurut Prasetyo (2012), sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar, sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*).

**Tabel 2.** Matriks Konfusi untuk Klasifikasi Dua Kelas

$f_{ij}$		Kelas hasil prediksi ( $j$ )	
		Kelas =1	Kelas = 0
Kelas asli ( $i$ )	Kelas =1	$f_{11}$	$f_{10}$
	Kelas = 0	$f_{01}$	$f_{00}$

Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi (Johnson dan Wichern, 2007). Pada penelitian ini digunakan pengukuran kinerja klasifikator berupa akurasi, *Apparent Error Rate* (APER), sensitivitas dan spesifisitas dengan penjelasan sebagai berikut :

### 2.9.1. Akurasi dan APER

Berdasarkan pada Tabel 2, untuk menghitung akurasi dan APER digunakan formula sebagai berikut :

$$\begin{aligned} \text{akurasi} &= \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \\ &= \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \end{aligned} \quad (12)$$

$$\begin{aligned} \text{APER} &= \frac{\text{jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} \\ &= \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}} \end{aligned} \quad (13)$$

### 2.9.3. Sensitivitas dan spesifisitas

Menurut Han dan Kamber (2006), ukuran sensitivitas dan spesifisitas dapat digunakan untuk mengetahui seberapa baik metode klasifikasi mengenali hasil yang positif dan seberapa baik dapat mengenali hasil yang negatif.

**Tabel 3.** Matriks Konfusi untuk Sensitivitas dan Spesifisitas

		Predicted class	
		$c_1$	$c_2$
Actual class	$c_1$	true positives (TP)	false negatives (FN)
	$c_2$	false positives (FP)	true negatives (TN)

Ukuran sensitivitas dan spesifisitas diperoleh sebagai berikut :

$$\text{sensitivitas} = \frac{TP}{TP + FN} \quad (14)$$



$$\text{spesifisitas} = \frac{TN}{FP + TN} \quad (15)$$

#### 2.9.4. Metode Hold-out

*Hold-out* adalah salah satu metode evaluasi klasifikator di mana set data (yang sudah diketahui label kelas aslinya) dipecah menjadi dua bagian terpisah, yaitu set data latih dan set data uji. Model klasifikasi kemudian dibangun berdasarkan set data latih dan kemudian kinerjanya diukur berdasarkan set data uji (Prasetyo, 2014).

### 3. METODE PENELITIAN

#### 3.1. Sumber Data dan Variabel Penelitian

Jenis data yang digunakan dalam penelitian ini merupakan data sekunder. Data yang dimaksud adalah data calon debitur tahun 2018 yang mengajukan permohonan KPR Multiguna *take over* ke Bank X di Provinsi DKI Jakarta terdiri atas hasil keputusan akhir kredit disetujui dan ditolak. Variabel dependen adalah hasil keputusan akhir kredit sedangkan variabel independen yang terpilih melalui uji independensi menggunakan statistik uji  $\chi^2$  (chi-square) dan digunakan dalam proses klasifikasi adalah variabel kepemilikan rumah, pekerjaan, pinjaman yang diajukan, serta penghasilan.

#### 3.2. Metode Analisis Data

Langkah dalam analisis data penelitian ini adalah :

1. Melakukan uji independensi variabel dependen dengan masing-masing variabel independen
2. Mendeskripsikan data
3. Mengelompokkan data ke dalam masing-masing kelas
4. Membagi data menjadi dua bagian yaitu sebagai data *training* dan data *testing*
5. Melakukan tahap penghitungan bobot yaitu menghitung nilai bobot untuk tiap atribut
6. Melakukan tahap klasifikasi data yaitu menghitung jarak (kedekatan lokasi) antara tiap data *testing* ke semua data *training*. Jarak yang digunakan adalah  $WSMC_{global}$ .
7. Mengurutkan nilai jarak  $WSMC_{global}$  dari yang terkecil hingga terbesar
8. Mengambil jarak  $WSMC_{global}$  terkecil sejumlah  $k$  tetangga terdekat sesuai dengan  $k$  yang dibutuhkan
9. Menetapkan kelas hasil dari proses klasifikasi. Kelas yang dipilih adalah kelas dengan nilai maksimum dari jumlah kelas yang muncul diantara sejumlah  $k$  tetangga terdekat
10. Menghitung APER hasil klasifikasi
11. Melakukan kembali langkah 6–10 hingga diperoleh nilai APER yang paling rendah dengan mengganti nilai  $k$

### 4. HASIL DAN PEMBAHASAN

#### 4.1. Uji Independensi antara Variabel Dependen dengan Variabel Independen

Langkah-langkah uji independensi menggunakan statistik uji  $\chi^2$  (chi-square) adalah sebagai berikut :

- a. Hipotesis
  - $H_0$  : Variabel X dan Variabel Y tidak berhubungan
  - $H_1$  : Variabel X dan Variabel Y berhubungan
- b. Taraf Signifikansi
  - $\alpha = 0,05$
- c. Kriteria Uji

$H_0$  ditolak jika  $\chi^2 > \chi^2_{tabel}$  atau Sig.  $< \alpha = 0,05$

d. Keputusan :

**Tabel 4.** Keputusan Uji  $\chi^2$  antara variabel independen (X) dengan variabel dependen (Y)

Variabel	$\chi^2$	db	$\chi^2_{tabel}$	Sig.	Keputusan
X <sub>1</sub>	1,278	3	7,815	0,734	H <sub>0</sub> diterima
X <sub>2</sub>	3,289	1	3,841	0,070	H <sub>0</sub> diterima
X <sub>3</sub>	237,264	2	5,991	0,004	H <sub>0</sub> ditolak
X <sub>4</sub>	0,086	1	3,841	0,769	H <sub>0</sub> diterima
X <sub>5</sub>	12,022	2	5,991	0,002	H <sub>0</sub> ditolak
X <sub>6</sub>	2,457	3	7,815	0,483	H <sub>0</sub> diterima
X <sub>7</sub>	2,670	3	7,815	0,445	H <sub>0</sub> diterima
X <sub>8</sub>	6,517	3	7,815	0,089	H <sub>0</sub> diterima
X <sub>9</sub>	8,563	3	7,815	0,036	H <sub>0</sub> ditolak
X <sub>10</sub>	12,203	2	5,991	0,002	H <sub>0</sub> ditolak

Artinya masing-masing variabel X<sub>3</sub>, X<sub>5</sub>, X<sub>9</sub>, dan X<sub>10</sub> berhubungan dengan variabel dependen (Y). Sehingga untuk selanjutnya variabel-variabel independen (X) yang digunakan dalam proses klasifikasi adalah variabel kepemilikan rumah, variabel pekerjaan, variabel pinjaman yang diajukan, dan variabel penghasilan.

#### 4.2. Metode *k* Nearest Neighbor dengan Pembobotan Global Gini Diversity Index

Pengolahan data menggunakan program MatLab R2015b dengan menjalankan *script* dari Wijaya (2017) yang telah dilakukan modifikasi oleh peneliti. Proporsi data *training* dan data *testing* dipilih berdasarkan nilai APER terkecil dari *output* program MatLab R2015b yang telah diujicobakan ke proporsi data *training* dan data *testing* sebesar 60:40, 70:30, 80:20 dan 90:10. Uji coba yang dilakukan dengan mengubah nilai *k*. Berdasarkan pengolahan yang telah dilakukan, terpilih nilai APER terkecil adalah 0,0798 yaitu pada proporsi 80:20 dengan *k*=7.

##### 1. Tahap Penghitungan Bobot

Berikut adalah hasil perhitungan bobot atribut ke-*d* atau dilambangkan  $\omega_d^{(GG)}$  dengan bantuan program MatLab R2015b yang dapat dilihat pada Tabel 5 sebagai berikut :

**Tabel 5.** Hasil  $\omega_d^{(GG)}$

Atribut ke- ( <i>d</i> )	$\omega_d^{(GG)}$
1	0,7357
2	0,7361
3	0,7355
4	0,7365

##### 2. Tahap Klasifikasi Data

Berikut adalah perhitungan jarak WSMC untuk data *testing* ke-1 :

- $WSMC_{global}(\mathbf{x}_1, \mathbf{x}_1, \varpi) = \sum_{d=1}^4 \omega_d \times I(x_{1d} \neq x_{1d}) = [(0,7357 \times 0) + (0,7361 \times 1) + (0,7355 \times 1) + (0,7365 \times 1)] = 2,2082$



- $WSMC_{global}(\mathbf{x}_2, \mathbf{x}_1, \varpi) = \sum_{d=1}^4 \omega_d \times I(x_{2d} \neq x_{1d}) = [(0,7357 \times 0) + (0,7361 \times 1) + (0,7355 \times 1) + (0,7365 \times 1)] = 2,2082$
- $WSMC_{global}(\mathbf{x}_3, \mathbf{x}_1, \varpi) = \sum_{d=1}^4 \omega_d \times I(x_{3d} \neq x_{1d}) = [(0,7357 \times 0) + (0,7361 \times 0) + (0,7355 \times 0) + (0,7365 \times 1)] = 0,7365$
- $WSMC_{global}(\mathbf{x}_4, \mathbf{x}_1, \varpi) = \sum_{d=1}^4 \omega_d \times I(x_{4d} \neq x_{1d}) = [(0,7357 \times 0) + (0,7361 \times 1) + (0,7355 \times 1) + (0,7365 \times 1)] = 2,2082$   
dan seterusnya sampai perhitungan jarak  $WSMC$  untuk data *testing* ke-1 dengan data *training* ke-1954 :
- $WSMC_{global}(\mathbf{x}_{1954}, \mathbf{x}_1, \varpi) = \sum_{d=1}^4 \omega_d \times I(x_{1954d} \neq x_{1d}) = [(0,7357 \times 0) + (0,7361 \times 1) + (0,7355 \times 0) + (0,7365 \times 1)] = 0,7361$

Hasil perhitungan jarak  $WSMC$  data *testing* ke-1 ke semua data *training* yang telah dilakukan pengurutan dari nilai terkecil hingga terbesar dapat dilihat pada Tabel 6 sebagai berikut :

**Tabel 6.** Jarak  $WSMC_{global}$

No.	Jarak $WSMC_{global}$
1	0
2	0
3	0
⋮	⋮
976	1,4720
977	1,4720
⋮	⋮
1952	2,9439
1953	2,9439
1954	2,9439

Berdasarkan proses klasifikasi yang telah dilakukan, diperoleh tabel matriks konfusi seperti terlihat pada Tabel 7 sebagai berikut :

**Tabel 7.** Matriks Konfusi Metode *k Nearest Neighbor* dengan Pembobotan *Global Gini Diversity Index*

Hasil Observasi ( <i>Actual Class</i> )	Hasil Prediksi ( <i>Predicted Class</i> )	
	Kelas 0 (disetujui)	Kelas 1 (ditolak)
Kelas 0 (disetujui)	447	1
Kelas 1 (ditolak)	38	3

Sehingga berdasarkan pada Tabel 7, diperoleh nilai  $APER$  dan akurasi sebagai berikut :

$$\begin{aligned}
 APER &= \frac{\text{jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} \\
 &= \frac{f_{01} + f_{10}}{(f_{00} + f_{01} + f_{10} + f_{11})} \\
 &= \frac{1 + 38}{(447 + 1 + 38 + 3)}
 \end{aligned}$$

$$= 0,0798$$

$$\begin{aligned} \text{Akurasi} &= (1 - \text{APER}) \times 100 \% \\ &= (1 - 0,0798) \times 100 \% \\ &= 92,02 \% \end{aligned}$$

Sementara itu, untuk nilai sensitivitas dan spesifisitas diperoleh sebagai berikut :

$$\begin{aligned} \text{Sensitivitas} &= \frac{TP}{TP + FN} \\ &= \frac{447}{447 + 1} \\ &= 0,9978 \end{aligned}$$

$$\begin{aligned} \text{Spesifisitas} &= \frac{TN}{FP + TN} \\ &= \frac{3}{38 + 3} \\ &= 0,0244 \end{aligned}$$

## 5. KESIMPULAN

Berdasarkan pada hasil dan pembahasan, atribut yang berpengaruh terhadap keputusan akhir kredit adalah atribut kepemilikan rumah, pekerjaan, pinjaman yang diajukan, dan penghasilan. Diperoleh nilai APER sebesar 0,0798 dan nilai akurasi sebesar 92,02 % pada proporsi hold-out 80:20 dengan k=7. Sementara itu untuk nilai sensitivitas atau proporsi yang teridentifikasi secara benar sebagai kelas 0 atau keputusan akhir kredit disetujui adalah sebesar 0,9978. Nilai spesifisitas atau proporsi yang teridentifikasi secara benar sebagai kelas 1 atau keputusan akhir kredit ditolak adalah sebesar 0,0244.

## DAFTAR PUSTAKA

- Ayodya, D. 2016. *Analisis Pengaruh Diferensiasi Produk, Kualitas Produk, Kualitas Layanan, dan Sikap Konsumen terhadap Keputusan Pengambilan KPR (Kredit Pemilikan Rumah)*. Diponegoro Journal of Management Vol. 4, No. 3.
- [BI] Bank Indonesia. 2009. *Hasil Kajian Kredit Konsumsi Mikro, Kecil, dan Menengah untuk Kegiatan Produktif*. Jakarta : BI.
- Chen, L., Guo G. 2015. *Nearest neighbor classification of categorical data by attributes weighting*. Expert Systems with Applications Vol. 42, No. 6 : Hal.3142–3149.
- Gorunescu, F. 2011. *Data Mining : Concepts, Model, and Techniques*. Jerman : Springer.
- Han, J., Kamber, M., 2006. *Data Mining : Concepts and Techniques*. San Francisco : Morgan Kaufmann Publishers.

- Johnson, R.A., Winchern, D.W. 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. New Jersey : Prentice Hall International, Inc.
- Kasmir. 2010. *Dasar-Dasar Perbankan*. Jakarta : PT. RAJAGRAFINDO PERSADA.
- [Kementerian PUPR] Kementerian Pekerjaan Umum dan Perumahan Rakyat. 2017. *Kamus Istilah Pembiayaan Perumahan*. Jakarta : Kementerian PUPR.
- Kuswati, S.M. 2015. *Cara Gampang Membeli Rumah tanpa Modal*. Publishing Langit.
- Prasetyo, Eko. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta : Penerbit ANDI Yogyakarta.
- Prasetyo, Eko. 2014. *Data Mining Mengolah Data menjadi Informasi Menggunakan MATLAB*. Yogyakarta : Penerbit ANDI Yogyakarta.
- Republik Indonesia. 1998. Undang-Undang Republik Indonesia Nomor 10 tahun 1998 tentang Perbankan. Jakarta.
- Tim Penulis Indira. 1996. *Kredit Pemilikan Rumah Tabel Pembayaran Hipotik*. Jakarta : PT Indira.
- Wijaya, R., Suciati, N., Khotimah, W.N. 2017. *Implementasi Nearest Neighbor pada Data Kategorik dengan Pembobotan Atribut menggunakan Weighted Simple Matching Coefficient*. Jurnal Teknik ITS Vol. 6, No. 2 : Hal. 2337-3520.