

## KLASIFIKASI RUMAH LAYAK HUNI DI KABUPATEN BREBES DENGAN MENGGUNAKAN METODE *LEARNING VECTOR QUANTIZATION* DAN NAIVE BAYES

Fitri Juniaty Simatupang<sup>1</sup>, Triastuti Wuryandari<sup>2</sup>, Suparti<sup>3</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

<sup>2,3</sup>Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro

<sup>1</sup>[fitrijuniatys@yahoo.co.id](mailto:fitrijuniatys@yahoo.co.id), <sup>2</sup>[triastuti@undip.ac.id](mailto:triastuti@undip.ac.id), <sup>3</sup>[supartisudargo@yahoo.co.id](mailto:supartisudargo@yahoo.co.id)

### ABSTRACT

House is a very basic need for everyone besides food and clothing. House can reflect the level of welfare and the level of health of its inhabitants. The advisability of a house as a good shelter can be seen from the structure and facilities of buildings. This research aims to analyze the classification of livable housing and determine the criteria of houses uninhabitable. The statistical method used are the Learning Vector Quantization and Naive Bayes. The data used in this final project are data of Survei Sosial Ekonomi Nasional (Susenans) Kor Keterangan Perumahan in 2014 Quarter 1 district of Kabupaten Brebes. In this research, the data divided into training data and testing data with the proportion that gives the highest accurate is 95% for training data and 5% for testing data. Training data will be used to generate the model and pattern formation, while testing data used to evaluate how accurate the model or pattern formed in classifying data through confusion tables. The results of analysis showed that the Learning Vector Quantization method gives 71,43% of classification accuracy, while Naive Bayes method gives 95,24% of classification accuracy. The Naive Bayes method has better classification accuracy than the Learning Vector Quantization method.

**Keywords:** House, Learning Vector Quantization, Naive Bayes, Classification

### 1. PENDAHULUAN

Salah satu indikator kesejahteraan rakyat adalah perumahan dan lingkungan. Dalam Undang-Undang Nomor 1 Tahun 2011 tentang Perumahan dan Kawasan Permukiman disebutkan dalam Pasal 1 Ayat 7 bahwa rumah adalah bangunan gedung yang berfungsi sebagai tempat tinggal yang layak huni, sarana pembinaan keluarga, cerminan harkat dan martabat penghuninya, serta aset bagi pemiliknya. Tempat tinggal atau rumah merupakan kebutuhan masyarakat yang sangat mendasar bagi setiap orang selain sandang dan pangan. Rumah yang layak huni dapat diketahui dari struktur bangunan rumah dan fasilitas rumah yang ada. Struktur bangunan rumah meliputi luas lantai, jenis lantai, jenis atap, dan jenis dinding. Sedangkan fasilitas rumah meliputi sumber air minum, sumber penerangan, dan tempat pembuangan akhir kotoran.

Dalam tulisan ini akan dilakukan pengklasifikasian rumah layak huni dan rumah tidak layak huni dengan melihat struktur bangunan rumah dan fasilitas rumah yang dimiliki setiap rumah tangga di Jawa Tengah, khususnya Kabupaten Brebes dengan metode klasifikasi yang digunakan adalah *Learning Vector Quantization* (LVQ) dan Naive Bayes. Tujuan dari tulisan ini adalah memperoleh hasil klasifikasi rumah layak huni di Kabupaten Brebes tahun 2014 dengan menggunakan metode LVQ dan Naive Bayes, memperoleh ketepatan klasifikasi rumah layak huni di Kabupaten Brebes dengan menggunakan metode LVQ dan Naive Bayes, dan memperoleh perbandingan ketepatan klasifikasi dengan metode LVQ dan Naive Bayes pada data Survei Sosial Ekonomi Nasional (Susenans) Kor Keterangan Perumahan Tahun 2014 Triwulan 1.

## 2. TINJAUAN PUSTAKA

### 2.1. Rumah

Rumah merupakan tempat berlindung terhadap gangguan dari luar serta tempat berkumpul dan berinteraksi bagi anggota rumah tangga [2]. Rumah dikatakan tidak layak huni apabila kondisi rumah memenuhi minimal salah satu kriteria dibawah ini [1]:

1. Luas lantai perkapita < 10 m<sup>2</sup>
2. Jenis lantai rumah adalah tanah
3. Jenis dinding rumah terbuat dari bambu/lainnya
4. Jenis atap rumah terbuat dari daun/lainnya
5. Sumber penerangan bukan listrik
6. Sumber air minum tidak layak
7. Tidak mempunyai fasilitas buang air besar
8. Tidak mempunyai tempat pembuangan akhir tinja berupa tangki septik

### 2.2. Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia [6]. Pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*) seperti berikut:

**Tabel 1.** Matriks Konfusi untuk Klasifikasi Dua Kelas

$f_{ij}$		Kelas hasil prediksi ( $j$ )	
		Kelas = 1	Kelas = 2
Kelas asli ( $i$ )	Kelas = 1	$f_{11}$	$f_{12}$
	Kelas = 2	$f_{21}$	$f_{22}$

Untuk menghitung akurasi dan laju error digunakan formula [6].

$$\text{Akurasi} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{11} + f_{22}}{f_{11} + f_{12} + f_{21} + f_{22}}$$

$$\text{Laju eror} = \frac{\text{jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{12} + f_{21}}{f_{11} + f_{12} + f_{21} + f_{22}}$$

### 2.3. Learning Vector Quantization

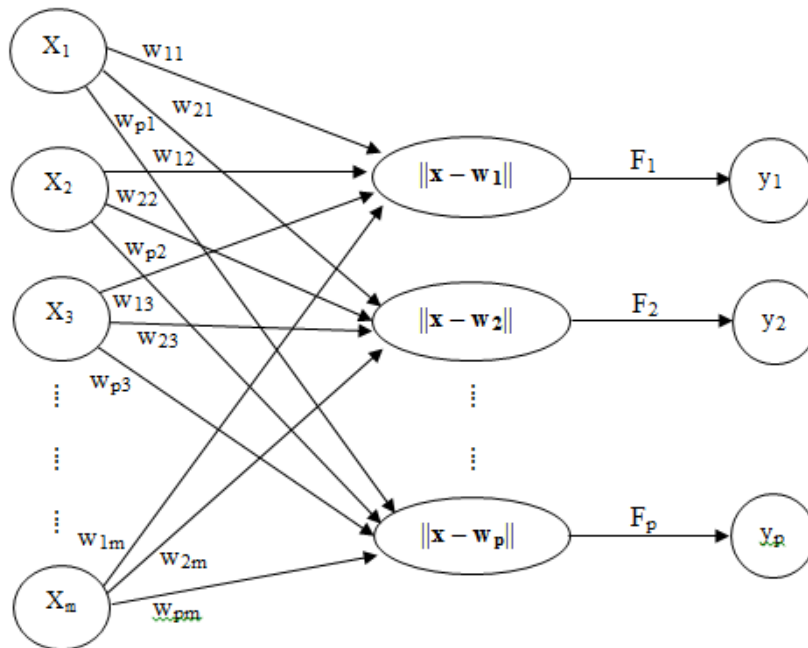
*Learning Vector Quantization* merupakan salah satu jenis *Artificial Neural Network* (ANN) yang berbasis *competitive learning* atau *winner take all*, dimana dari nilai keluaran yang diberikan neuron dalam layer keluaran hanya neuron pemenang (neuron yang mempunyai nilai terkecil) saja yang diperhatikan [7]. *Learning Vector Quantization* (LVQ) merupakan suatu metode untuk melakukan pelatihan terhadap lapisan-lapisan kompetitif yang terawasi [10]. Lapisan kompetitif akan belajar secara otomatis untuk melakukan klasifikasi terhadap vektor *input* yang diberikan. Apabila beberapa vektor *input* memiliki jarak yang sangat berdekatan, maka vektor-vektor *input* tersebut dikelompokkan ke dalam kelas yang sama. Pengukuran jarak yang digunakan adalah jarak *Euclidean* dengan formula [6]:

$$d(\mathbf{x}_i, \mathbf{w}_k) = \|\mathbf{x}_i - \mathbf{w}_k\| = \sqrt{\sum_{j=1}^m (x_{ij} - w_{kj})^2} \quad (1)$$

dengan  $i = 1, 2, \dots, n$  merupakan banyaknya data,  $j = 1, 2, \dots, m$  merupakan banyaknya variabel data,  $k = 1, 2, \dots, p$  merupakan banyaknya bobot.

Koneksi antar neuron tersebut dihubungkan dengan bobot/*weight*. Bobot merupakan nilai matematis dari koneksi yang mentransfer data dari satu lapisan ke lapisan lainnya,

yang berfungsi untuk mengatur jaringan sehingga dapat menghasilkan *output* yang diinginkan sekaligus bertujuan membuat jaringan itu belajar [10]. Bobot yang digunakan harus mewakili masing-masing kelas klasifikasi karena vektor bobot tersebut berfungsi untuk menghubungkan setiap neuron pada lapisan *input* dengan masing-masing neuron pada lapisan *output*. Vektor bobot biasanya dituliskan dengan  $w_{kj} = (w_{k1}, w_{k2}, w_{k3}, \dots, w_{km})$  dimana  $k$  merupakan banyaknya kelas,  $m$  banyaknya variabel yang digunakan. Pola-pola akan disajikan dalam bentuk vektor. Model jaringan LVQ secara umum dapat dilihat seperti Gambar 1.



**Gambar 1.** Arsitektur Jaringan LVQ

Parameter-parameter yang digunakan pada metode LVQ ini adalah sebagai berikut [10]:

1. Alfa (*Learning Rate*)  
Alfa didefinisikan sebagai tingkat pembelajaran. Jika alfa terlalu besar, maka algoritma akan menjadi tidak stabil. Sebaliknya jika alfa terlalu kecil, maka prosesnya akan terlalu lama. Nilai alfa adalah  $0 < \alpha < 1$ .
2. DecAlfa (Penurunan *Learning Rate*)  
Yaitu penurunan tingkat pembelajaran.
3. MinAlfa (Minimum *Learning Rate*)  
Yaitu minimal nilai tingkat pembelajaran yang masih diperbolehkan.
4. MaxEpoch (Maksimum epoch)  
Yaitu jumlah epoch atau iterasi maksimum yang boleh dilakukan selama pelatihan. Iterasi akan diberhentikan jika nilai epoch melebihi epoch maksimum.

Proses pembelajaran metode LVQ dikerjakan dengan algoritma sebagai berikut [5]:

0. Tetapkan:
  - a. Bobot awal variabel *input* ke- $j$  menuju kelas ke- $k$  :  $w_{kj}$ , dengan  $k = 1, 2, \dots, p$  dan  $j = 1, 2, \dots, m$ .
  - b. Maksimum Epoch: MaxEpoch.
  - c. Parameter *learning rate*:  $\alpha$ .
  - d. Pengurangan *learning rate*: Dec $\alpha$ .
  - e. Minimal *learning rate* yang diperbolehkan: Min $\alpha$

1. Masukkan:
  - a. Data *input*:  $X_{ij}$ ;  
dengan  $i = 1, 2, \dots, n$ ; dan  $j = 1, 2, \dots, m$ .
  - b. Target berupa kelas:  $T_i$ ;  
dengan  $i = 1, 2, \dots, n$ .
2. Tetapkan kondisi awal:  $\text{epoch} = 0$ ;
3. Kerjakan jika: ( $\text{epoch} \leq \text{MaxEpoch}$ ) dan ( $\alpha \geq \text{Min}\alpha$ )
  - a.  $\text{Epoch} = \text{epoch} + 1$ ;
  - b. Kerjakan untuk  $i = 1$  sampai  $n$ 
    - i. Tentukan  $k$  sedemikian hingga  $\|\mathbf{x}_i - \mathbf{w}_k\|$  minimum;  
dengan  $k = 1, 2, \dots, p$
    - ii. Perbaiki  $\mathbf{w}_j$  dengan ketentuan:
      - Jika  $T = C_j$  maka:  
$$\mathbf{w}_k^{\text{baru}} = \mathbf{w}_k^{\text{lama}} + \alpha (\mathbf{x}_i - \mathbf{w}_k^{\text{lama}}) \quad (2)$$
      - Jika  $T \neq C_j$  maka:  
$$\mathbf{w}_k^{\text{baru}} = \mathbf{w}_k^{\text{lama}} - \alpha (\mathbf{x}_i - \mathbf{w}_k^{\text{lama}}) \quad (3)$$
  - c. Kurangi nilai  $\alpha$ .  
Pengurangan  $\alpha$  bisa dilakukan dengan menggunakan rumus:  
$$\alpha = \alpha - \text{Dec}\alpha * \alpha \quad (4)$$

Algoritma pengujian pada metode LVQ adalah sebagai berikut [5]:

1. Masukkan data yang akan diuji, misal:  $x_{ij}$ ; dengan  $i = 1, 2, \dots, n$ ; dan  $j = 1, 2, \dots, m$ .
2. Kerjakan untuk  $i = 1$  sampai  $n$ 
  - a. Tentukan  $k$  sedemikian hingga  $\|\mathbf{x}_i - \mathbf{w}_k\|$  minimum;  
dengan  $k = 1, 2, \dots, p$ .
  - b.  $k$  adalah kelas untuk  $x_i$ .

#### 2.4. Naive Bayes Classifier

Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat [6]. *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output* [8]. Pengklasifikasian Naive Bayes dilakukan dengan memilih probabilitas akhir (posterior) tertinggi dari masing-masing kelas. Formulasi Naive Bayes untuk klasifikasi atau menghitung probabilitas posterior adalah sebagai berikut:

$$P(Y | X) = \frac{P(Y) \prod_{j=1}^m P(X_j | Y)}{P(X)} \quad (5)$$

dengan:

- a.  $P(Y | X)$  : probabilitas data dengan vektor  $X$  pada kelas  $Y$ .
- b.  $P(Y)$  : probabilitas awal kelas  $Y$ .
- c.  $\prod_{j=1}^m P(X_j | Y)$  : probabilitas independen kelas  $Y$  dari semua fitur dalam vektor  $X$ .
- d.  $P(X)$  : probabilitas awal dari  $X$ .

Probabilitas  $P(X)$  selalu tetap sehingga dalam perhitungan prediksi nantinya dapat dihilangkan dan hanya menghitung bagian  $P(Y) \prod_{j=1}^m P(X_j | Y)$  saja dengan memilih nilai yang terbesar sebagai kelas yang dipilih untuk hasil prediksi.

Untuk fitur bertipe kategoris, seperti data “jenis kelamin” dengan nilai {pria, wanita}, nilai probabilitas prior (awal) pada Bayes dihitung dengan menggunakan rumus:

$$P(X_i = x_i | Y = y_j) = \frac{\sum_{i=1}^n X_{ij}}{n} \quad (6)$$

dimana  $P(X_i = x_i | Y = y_j)$  merupakan probabilitas variabel pada kelas  $y_j$ ,  $\sum_{i=1}^n X_{ij}$  merupakan jumlah data latih pada kelas  $j$ ,  $n$  merupakan banyaknya data kelas  $j$ . Sedangkan untuk fitur bertipe numerik (kontinu) ada perlakuan khusus sebelum diproses menggunakan metode Naive Bayes. Caranya adalah:

- Transformasi fitur kontinu ke dalam fitur ordinal.
- Asumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data latih. Distribusi Gaussian biasanya dipilih untuk mempresentasikan probabilitas bersyarat dari fitur kontinu pada sebuah kelas  $P(x_j|y)$ , sedangkan distribusi Gaussian dikarakteristikkan dengan dua parameter yaitu mean ( $\mu$ ) dan varian ( $\sigma^2$ ). Untuk menghitung nilai mean ( $\mu$ ) digunakan rumus:

$$\mu_k = \frac{\sum_{i=1}^n X_{ik}}{n} \quad (7)$$

dimana  $\mu_k$  merupakan nilai rata-rata data latih kelas  $k$ ,  $\sum_{i=1}^n x_{ik}$  merupakan jumlah data latih pada kelas  $k$ ,  $n$  merupakan banyaknya data latih kelas  $k$ . Sedangkan untuk menghitung nilai varian ( $\sigma^2$ ) digunakan rumus:

$$\sigma_k^2 = \frac{\sum_{i=1}^n (X_{ik} - \mu_k)^2}{n-1} \quad (8)$$

dimana  $\sigma_k^2$  merupakan nilai varian data latih kelas  $k$ ,  $\sum_{i=1}^n (x_{ik} - \mu_k)$  merupakan jumlah dari selisih data latih pada kelas  $k$  dengan rata-rata data latih kelas  $k$ ,  $n$  merupakan banyaknya data latih kelas  $k$ .

Untuk setiap kelas  $y_k$ , probabilitas untuk fitur  $x_i$  pada distribusi Gaussian adalah:

$$P(X = x_i | Y = y_k) = f(x_i; \mu_k, \sigma_k) \quad (9)$$

$$f(x_i; \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp \left[ -\frac{(x_i - \mu_k)^2}{2\sigma_k^2} \right] \quad (10)$$

Persamaan (9) tidak sepenuhnya benar: probabilitas dengan variabel random kontinu memiliki nilai sama dengan nol. Probabilitas untuk variabel random kontinu adalah

$$P(x \leq X \leq x+\Delta) = \int_x^{x+\Delta} f(x; \mu, \sigma) dx . \text{ Dengan definisi dari turunan, } \lim_{\Delta \rightarrow 0} \frac{P(x \leq X \leq x+\Delta)}{\Delta} = f(x; \mu, \sigma). \text{ Jadi, untuk } \Delta \text{ yang sangat kecil } P(X = x) \approx f(x; \mu, \sigma)\Delta.$$

Persamaan  $P(C = c | X = x) = \frac{P(C=c)P(X=x|C=c)}{P(X=x)}$  untuk masing-masing kelas tidak berlaku setelah dilakukan normalisasi, sehingga Persamaan (9) dapat digunakan [4].

Klasifikasi dengan Naive Bayes bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik Naive Bayes sebagai berikut [6]:

- Metode Naive Bayes teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outlier*).
- Tangguh terhadap atribut yang tidak relevan.
- Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi Naive Bayes.

### 3. METODE PENELITIAN

#### 3.1. Jenis dan Sumber Data

Data yang digunakan dalam tulisan ini berupa data sekunder yaitu data Survei Sosial Ekonomi Nasional (Susenas) Kor Keterangan Perumahan Tahun 2014 Triwulan 1 yang diambil dari Badan Pusat Statistika (BPS) Jawa Tengah untuk Kabupaten Brebes.

### 3.2. Populasi dan Sampel

Jumlah rumah tangga di Kabupaten Brebes sebanyak 453.944 rumah tangga [3]. Keseluruhan rumah tangga dibagikan ke dalam 4 triwulan, sehingga populasi yang digunakan dalam penelitian ini sebesar 113.486 rumah tangga. Sedangkan untuk pemilihan sampelnya ditentukan dari BPS Pusat dengan menggunakan sampel acak. Penentuan ukuran sampel minimum dalam penelitian ini menggunakan rumus Slovin. Ukuran sampel dengan *slovin* dapat dihitung dengan rumus [9]:

$$n = \frac{N}{1 + (N \times e^2)}$$

dimana:  $n$  : ukuran sampel minimum  
 $N$  : jumlah populasi  
 $e$  : tingkat kesalahan yang digunakan dalam penelitian

Sehingga ukuran sampel minimum yang digunakan dalam penelitian ini sebesar:

$$n = \frac{113486}{1 + (113486 \times (0,05)^2)} = 398,59 \approx 399 \text{ rumah tangga}$$

Ukuran sampel minimum data terpenuhi karena jumlah keseluruhan data yang digunakan sebanyak 838 data.

### 3.3. Variabel Data

Variabel yang digunakan dibagi menjadi 2 yaitu variabel respon dan variabel prediktor.

1. Variabel respon:  
Y = Klasifikasi rumah
2. Variabel-variabel prediktor:  
X1 = Jenis atap rumah  
X2 = Jenis dinding rumah  
X3 = Jenis lantai rumah  
X4 = Luas lantai per kapita  
X5 = Sumber air minum/bersih  
X6 = Penggunaan fasilitas dan jenis tempat buang air besar

### 3.4. Teknik Pengolahan Data

Data diolah dengan menggunakan metode LVQ dan Naive Bayes dengan menggunakan Software *Excel* dan *Matlab*. Adapun langkah-langkah yang dilakukan adalah sebagai berikut:

1. Membagi data menjadi 2, yaitu data latih dan data uji dengan 4 percobaan proporsi pembagian data yaitu 95% data latih dan 5% data uji, 90% data latih dan 10% data uji, 80% data latih dan 20% data uji, 70% data latih dan 30% data uji.
2. Melakukan klasifikasi menggunakan metode LVQ dengan langkah-langkah:
  - Menentukan bobot awal, MaxEpoch,  $\alpha$ , Dec $\alpha$ , Min $\alpha$
  - Menginput data latih dan kondisi awal epoch
  - Menghitung epoch dan jarak setiap data ke setiap bobot
  - Menentukan jarak minimum dan kelas data
  - Memperbaharui bobot
  - Mengurangi nilai  $\alpha$
  - Jika epoch  $\leq$  MaxEpoch dan  $\alpha \geq$  Min $\alpha$ , maka dilakukan langkah ke-4 sampai ke-8
  - Menentukan bobot akhir dan menginput data uji

- Menghitung jarak setiap data ke setiap bobot
  - Menentukan jarak minimum dan kelas data
  - Mengevaluasi hasil klasifikasi
3. Melakukan klasifikasi menggunakan metode Naive Bayes dengan langkah-langkah:
- Menginput data latih
  - Menghitung probabilitas prior ( $P(Y)$ ) dari data latih yang bertipe kategorik
  - Menghitung rata-rata dan varian sampel dari data bertipe numeric
  - Menginput data uji
  - Menghitung probabilitas atribut terhadap masing-masing kelas ( $P(X_j|Y)$ ) pada data uji yang bertipe kategorik berdasarkan data latih
  - Menghitung probabilitas atribut terhadap masing-masing kelas ( $P(X_j|Y)$ ) pada data uji yang bertipe numeric berdasarkan distribusi Gaussian dengan rata-rata dan varian sampel pada data latih
  - Menghitung perkalian probabilitas dengan probabilitas atribut pada masing-masing kelas ( $P(Y) P(X_j|Y)$ ).
  - Mencari nilai maksimal dari ( $P(Y) P(X_j|Y)$ ) pada kedua kelas.
  - Nilai terbesar dari penghitungan merupakan hasil prediksi.
  - Mengevaluasi hasil klasifikasi.
4. Membandingkan ketepatan klasifikasi antara metode LVQ dan Naive Bayes.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Deskripsi Data

Data dideskriptifkan seperti pada Tabel 2.

**Tabel 2.** Deskripsi Variabel Respon (Y)

Status	Frekuensi (n)	Persentase (%)
Tidak layak huni	372	44,39
Layak huni	466	55,61
<b>Total</b>	<b>838</b>	<b>100,00</b>

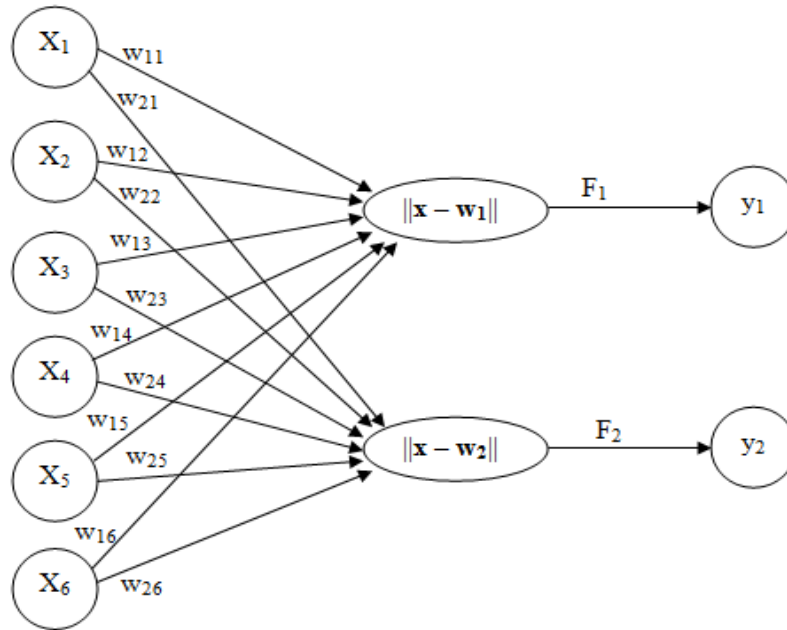
Data dibagi menjadi dua kelompok, yaitu data latih dan data uji. Proporsi pembagiannya telah dilakukan beberapa percobaan, namun dari keempat percobaan yang telah dilakukan diperoleh nilai ketepatan klasifikasi tertinggi pada data dengan proporsi data latih 95% dan data uji 5% yaitu sebesar 95,24%. Adapun pembagiannya sebagai berikut:

$$n_{\text{latih}} = \left(\frac{95}{100} \times 372\right) + \left(\frac{95}{100} \times 466\right) = 353,4 + 442,7 \approx 353 + 443 = 796$$

$$n_{\text{uji}} = \left(\frac{5}{100} \times 372\right) + \left(\frac{5}{100} \times 466\right) = 18,6 + 23,3 \approx 19 + 23 = 42$$

### 4.2. Analisis *Learning Vector Quantization* (LVQ)

Arsitektur jaringan LVQ pada penelitian ini dapat dilihat pada Gambar 2.



**Gambar 2.** Arsitektur LVQ pada Data

Bobot awal yang diambil secara acak dari data yaitu  $\mathbf{w}_1 = (2; 1; 1; 18; 6; 3)$  dan  $\mathbf{w}_2 = (2; 1; 3; 34,67; 5; 3)$ . Parameter-parameter yang digunakan adalah  $\alpha = 0,1$ ;  $Dec\alpha = 0,2$ ;  $Min\alpha = 0,001$ ;  $MaxEpoch = 50$ . Proses pelatihan dimulai dari epoch yang pertama, data latih yang pertama  $\mathbf{x}_1 = (x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}) = (2; 1; 1; 24; 6; 3)$ . Penghitungan jarak dilakukan dengan menggunakan rumus pada Persamaan (1), diperoleh jarak *euclidean* data latih pertama dengan kelas pertama sebagai berikut:

$$\begin{aligned} d(\mathbf{x}_1, \mathbf{w}_1) &= \|\mathbf{x}_1 - \mathbf{w}_1\| = \sqrt{(x_{11} - w_{11})^2 + (x_{12} - w_{12})^2 + \dots + (x_{16} - w_{16})^2} \\ &= \sqrt{(2 - 2)^2 + (1 - 1)^2 + (1 - 1)^2 + (24,00 - 18,00)^2 + (6 - 6)^2 + (3 - 3)^2} \\ &= 6 \end{aligned}$$

Sedangkan jarak *euclidean* data latih pertama dengan kelas kedua sebagai berikut:

$$\begin{aligned} d(\mathbf{x}_1, \mathbf{w}_2) &= \|\mathbf{x}_1 - \mathbf{w}_2\| = \sqrt{(x_{11} - w_{21})^2 + (x_{12} - w_{22})^2 + \dots + (x_{16} - w_{26})^2} \\ &= \sqrt{(2 - 2)^2 + (1 - 1)^2 + (1 - 3)^2 + (24,00 - 34,67)^2 + (6 - 5)^2 + (3 - 3)^2} \\ &= 10,902 \end{aligned}$$

Data latih ke-1 diklasifikasikan ke dalam kelas 1 karena  $\|\mathbf{x}_1 - \mathbf{w}_2\| > \|\mathbf{x}_1 - \mathbf{w}_1\|$ . Selanjutnya nilai bobot akan diperbaharui menggunakan Persamaan (2), jika kelas hasil klasifikasi sama dengan kelas asli. Jika kelas hasil klasifikasi tidak sama dengan kelas asli, digunakan Persamaan (3). Data latih pertama memiliki kelas asli 1 dan kelas hasil klasifikasi yang diperoleh adalah kelas 1, sehingga bobot pertama diperbaharui menggunakan Persamaan (2). Nilai vektor bobot pertama diperoleh:

$$\begin{aligned} \mathbf{w}_1^{\text{baru}} &= (\mathbf{w}_1^{\text{lama}})' + \alpha ((\mathbf{x}_1)' - (\mathbf{w}_1^{\text{lama}})') \\ \mathbf{w}_1^{\text{baru}} &= \begin{pmatrix} 2 \\ 1 \\ 1 \\ 18,00 \\ 6 \\ 3 \end{pmatrix} + (0,1) \left\{ \begin{pmatrix} 2 \\ 1 \\ 1 \\ 24,00 \\ 6 \\ 3 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \\ 1 \\ 18,00 \\ 6 \\ 3 \end{pmatrix} \right\} = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 18,6 \\ 6 \\ 3 \end{pmatrix} \end{aligned}$$

dimana:

- $\mathbf{w}_1^{\text{baru}}$  : nilai vektor bobot kelas 1 yang sudah diperbaharui
- $\mathbf{w}_1^{\text{lama}}$  : nilai vektor bobot kelas 1 yang belum diperbaharui
- $\mathbf{x}_1$  : vektor data latih ke-1

Dari perhitungan diatas diperoleh vektor bobot yang menghubungkan input dengan kelas ke-1 yaitu  $\mathbf{w}_1^{\text{baru}} = (2; 1; 1; 18,6; 6; 3)$ .



Setelah diperoleh bobot baru, proses pelatihan dilanjutkan pada data latih ke-2 sampai data ke-794 dengan langkah-langkah yang sama. Proses tersebut masih dilakukan pada epoch pertama. Untuk epoch selanjutnya akan terlebih dahulu dilakukan pengurangan nilai  $\alpha$  (*learning rate*) seperti berikut:

$$\begin{aligned}\alpha_{\text{baru}} &= \alpha_{\text{lama}} - \text{Deca} * \alpha_{\text{lama}} \\ &= 0,1 - (0,2) * (0,1) \\ &= 0,1 - 0,02 \\ &= 0,08\end{aligned}$$

Proses pelatihan akan dilakukan lagi pada data latih ke-1 sampai data latih ke-794 dengan menggunakan  $\alpha = 0,08$ . Proses dilakukan sampai epoch mencapai maksimal. Setelah itu proses pelatihan akan berhenti dan diperoleh nilai bobot akhir sebagai berikut sekaligus yang akan digunakan pada proses pengujian:

$$\mathbf{w}_1 = (2,2377; 4,2825; 10,2673; 0,6282; 10,2505; 10,0105)$$

$$\mathbf{w}_2 = (1,9321; 0,8307; 0,4932; 42,3547; 3,6520; 0,0054)$$

Proses pelatihan berhenti pada epoch ke-21 dengan nilai  $\alpha = 0,0011529$ . Selanjutnya pengklasifikasian data uji dilakukan sebagai berikut:

Data uji ke-1 =  $\mathbf{x}_1 = (2; 1; 1; 13,75; 6; 3)$

Penghitungan jarak *euclidean* data uji ke-1 dengan kelas pertama sebagai berikut:

$$\begin{aligned}d(\mathbf{x}_1, \mathbf{w}_1) &= \|\mathbf{x}_1 - \mathbf{w}_1\| = \sqrt{(x_{11} - w_{11})^2 + (x_{12} - w_{12})^2 + \dots + (x_{16} - w_{16})^2} \\ &= \sqrt{(2 - 2,2377)^2 + (1 - 4,2825)^2 + \dots + (3 - 10,0105)^2} \\ &= 18,3333\end{aligned}$$

Sedangkan jarak *euclidean* data latih ke-1 dengan kelas kedua sebagai berikut:

$$\begin{aligned}d(\mathbf{x}_1, \mathbf{w}_2) &= \|\mathbf{x}_1 - \mathbf{w}_2\| = \sqrt{(x_{11} - w_{21})^2 + (x_{12} - w_{22})^2 + \dots + (x_{16} - w_{26})^2} \\ &= \sqrt{(2 - 1,9321)^2 + (1 - 0,8307)^2 + \dots + (3 - 0,0054)^2} \\ &= 28,8617\end{aligned}$$

Hasil yang diperoleh  $\|\mathbf{x}_1 - \mathbf{w}_2\| > \|\mathbf{x}_1 - \mathbf{w}_1\|$ , sehingga data uji ke-1 diklasifikasikan pada kelas 1. Proses klasifikasi dilakukan hingga data uji ke-42 dengan menggunakan langkah-langkah yang sama. Ketepatan hasil klasifikasi metode LVQ ditampilkan seperti berikut:

**Tabel 3.** Ketepatan Hasil Klasifikasi *Learning Vector Quantization*

Kelas Asli	Kelas Hasil Klasifikasi	
	Kelas 1	Kelas 2
Kelas 1	13	6
Kelas 2	6	17

Berdasarkan tabel dapat dihitung nilai akurasi dan laju error sebagai berikut:

$$\begin{aligned}\text{Akurasi} &= \frac{\text{jumlah data yang diklasifikasikan secara benar}}{\text{jumlah semua data yang diklasifikasikan}} \\ &= \frac{f_{11} + f_{22}}{f_{11} + f_{12} + f_{21} + f_{22}} = \frac{13 + 17}{13 + 6 + 6 + 17} = \frac{30}{42} = 0,7143\end{aligned}$$

$$\begin{aligned}\text{Laju error} &= \frac{\text{jumlah data yang diklasifikasikan secara salah}}{\text{jumlah semua data yang diklasifikasikan}} \\ &= \frac{f_{12} + f_{21}}{f_{11} + f_{12} + f_{21} + f_{22}} = \frac{6 + 6}{13 + 6 + 6 + 17} = \frac{12}{42} = 0,2857\end{aligned}$$

Ketepatan klasifikasi metode LVQ sebesar 0,7143 atau 71,43% dengan laju error sebesar 0,2857 atau 28,57%.

### 4.3. Analisis Naive Bayes Classifier

Proses pelatihan dilakukan dengan menghitung nilai probabilitas prior masing-masing variabel. Untuk variabel respon (y) dengan tipe data kategorik, penghitungan probabilitas prior dilakukan menggunakan Persamaan (6), sehingga diperoleh seperti pada Tabel 4.

**Tabel 4.** Probabilitas Prior Variabel Respon

	<b>Frekuensi</b>	<b>Probabilitas Prior</b>
<b>Tidak Layak</b>	352	0,443324937
<b>Layak</b>	442	0,556675063
<b>Total</b>	794	

Probabilitas prior dari variabel jenis atap rumah ( $X_1$ ):

**Tabel 5.** Probabilitas Prior Variabel Jenis Atap Rumah ( $X_1$ )

	<b>Frekuensi <math>X_1</math></b>		<b>Probabilitas <math>X_1</math></b>	
	<b>Tidak Layak</b>	<b>Layak</b>	<b>Tidak Layak</b>	<b>Layak</b>
<b>Beton</b>	6	11	0,017045455	0,024886878
<b>Genteng</b>	329	422	0,934659091	0,954751131
<b>Sirap</b>	0	0	0	0
<b>Seng</b>	10	5	0,028409091	0,011312217
<b>Asbes</b>	7	4	0,019886364	0,009049774
<b>Ijuk/rumbia</b>	0	0	0	0
<b>Lainnya</b>	0	0	0	0
<b>Total</b>	352	442		

Probabilitas prior dari variabel jenis dinding rumah ( $X_2$ ):

**Tabel 6.** Probabilitas Prior Variabel Jenis Dinding Rumah ( $X_2$ )

	<b>Frekuensi <math>X_2</math></b>		<b>Probabilitas <math>X_2</math></b>	
	<b>Tidak Layak</b>	<b>Layak</b>	<b>Tidak Layak</b>	<b>Layak</b>
<b>Tembok</b>	269	431	0,764204545	0,975113122
<b>Kayu</b>	20	11	0,056818182	0,024886878
<b>Bambu</b>	55	0	0,15625	0
<b>Lainnya</b>	8	0	0,022727273	0
<b>Total</b>	352	442		

Probabilitas prior dari variabel jenis lantai rumah ( $X_3$ ):

**Tabel 7.** Probabilitas Prior Variabel Jenis Lantai Rumah ( $X_3$ )

	<b>Frekuensi <math>X_3</math></b>		<b>Probabilitas <math>X_3</math></b>	
	<b>Tidak Layak</b>	<b>Layak</b>	<b>Tidak Layak</b>	<b>Layak</b>
<b>Marmer/Keramik/Granit</b>	119	288	0,338068182	0,65158371
<b>Tegel/teraso</b>	21	49	0,059659091	0,110859729
<b>Semen</b>	93	105	0,264204545	0,237556561
<b>Kayu</b>	1	0	0,002840909	0
<b>Tanah</b>	114	0	0,323863636	0
<b>Lainnya</b>	4	0	0,011363636	0
<b>Total</b>	352	442		

Probabilitas prior dari variabel sumber air minum ( $X_5$ ):

**Tabel 8.** Probabilitas Prior Variabel Sumber Air Minum ( $X_5$ )

	<b>Frekuensi <math>X_5</math></b>		<b>Probabilitas <math>X_5</math></b>	
	<b>Tidak Layak</b>	<b>Layak</b>	<b>Tidak Layak</b>	<b>Layak</b>
<b>Air kemasan bermerk</b>	3	26	0,008522727	0,058823529
<b>Air isi ulang</b>	32	69	0,090909091	0,156108597
<b>Leding meteran</b>	6	34	0,017045455	0,076923077

Leding eceran	53	81	0,150568182	0,183257919
Sumur bor/pompa	53	82	0,150568182	0,185520362
Sumur terlindung	133	120	0,377840909	0,271493213
Sumur tak terlindung	6	0	0,017045455	0
Mata air terlindung	60	30	0,170454545	0,067873303
Mata air tak terlindung	4	0	0,011363636	0
Air sungai	2	0	0,005681818	0
Air hujan	0	0	0	0
Lainnya	0	0	0	0
<b>Total</b>	<b>352</b>	<b>442</b>		

Probabilitas prior dari variabel fasilitas dan tempat buang air besar ( $X_6$ ):

**Tabel 9.** Probabilitas Prior Variabel Fasilitas dan Tempat BAB ( $X_6$ )

	Frekuensi $X_6$		Probabilitas $X_6$	
	Tidak Layak	Layak	Tidak Layak	Layak
Sendiri-Tangki/SPAL	64	405	0,181818182	0,916289593
Bersama-Tangki/SPAL	19	37	0,053977273	0,083710407
Lainnya	269	0	0,764204545	0
<b>Total</b>	<b>352</b>	<b>442</b>		

Untuk variabel dengan tipe data numerik, penghitungan probabilitas dilakukan dengan Distribusi Gaussian. Dari data latih dengan menggunakan rumus pada Persamaan (8) dan (9), diperoleh nilai mean dan varian masing-masing kelas seperti pada Tabel 10.

**Tabel 10.** Nilai Mean dan Varian Variabel Luas Lantai per Kapita ( $X_4$ )

	$X_4$	
	Tidak Layak	Layak
<b>Rata-rata</b>	21,91	31,22
<b>Variansi</b>	317,93	508,15

Sedangkan untuk nilai probabilitas prior dari variabel luas lantai per kapita langsung diterapkan pada data uji, sehingga diperoleh hasil sebagai berikut:

Rumah dengan luas lantai per kapita 13,75 m<sup>2</sup> memiliki nilai probabilitas prior untuk masing-masing kelas sebagai berikut:

$$\begin{aligned}
 P(X_4 = 13,75 \mid \text{tidak layak}) &= \frac{1}{\sqrt{2\pi\sigma_{\text{tidak layak}}^2}} \exp \left[ -\frac{(x_4 - \mu_{\text{tidak layak}})^2}{2\sigma_{\text{tidak layak}}^2} \right] \\
 &= \frac{1}{\sqrt{2\pi(317,93)}} \exp \left[ -\frac{(13,75 - 21,91)^2}{2(317,93)} \right] = 0,020151666
 \end{aligned}$$

$$\begin{aligned}
 P(X_4 = 13,75 \mid \text{layak}) &= \frac{1}{\sqrt{2\pi\sigma_{\text{layak}}^2}} \exp \left[ -\frac{(x_4 - \mu_{\text{layak}})^2}{2\sigma_{\text{layak}}^2} \right] \\
 &= \frac{1}{\sqrt{2\pi(508,15)}} \exp \left[ -\frac{(13,75 - 31,22)^2}{2(508,15)} \right] = 0,013107709
 \end{aligned}$$

Hal yang sama dilakukan pada data uji ke-2 sampe ke-42. Selanjutnya dihitung nilai probabilitas independen dengan mengalikan semua probabilitas prior masing-masing kelas sesuai ketentuan variabel data ujinya. Nilai probabilitas bersyarat data uji ke-1 untuk masing-masing kelas adalah sebagai berikut:

$$\begin{aligned}
P(X | \text{tidak layak}) &= P(X_1 = \text{Genteng} | \text{tidak layak}) \times P(X_2 = \text{Tembok} | \text{tidak layak}) \times \\
&P(X_3 = \text{Marmer/keramik/granit} | \text{tidak layak}) \times P(X_4 = 13,75 | \text{tidak layak}) \times P(X_5 = \text{Sumur terlindung} | \text{tidak layak}) \times P(X_6 = \text{Lainnya} | \text{tidak layak}) \\
&= 0,934659091 \times 0,764204545 \times 0,338068182 \times 0,020151666 \times 0,377840909 \times 0,764204545 = 0,001405066 \\
P(X | \text{layak}) &= P(X_1 = \text{Genteng} | \text{layak}) \times P(X_2 = \text{Tembok} | \text{layak}) \times P(X_3 = \text{Marmer/keramik/granit} | \text{layak}) \times P(X_4 = 13,75 | \text{layak}) \times P(X_5 = \text{Sumur terlindung} | \text{layak}) \times P(X_6 = \text{Lainnya} | \text{layak}) \\
&= 0,954751131 \times 0,975113122 \times 0,65158371 \times 0,01310771 \times 0,271493213 \times 0 = 0
\end{aligned}$$

Perhitungan dilanjutkan sampai data ke-42. Selanjutnya dilakukan perhitungan nilai probabilitas akhir atau posterior dari data uji menggunakan rumus  $P(Y) \prod_{j=1}^m P(X_j | Y)$ . Probabilitas posterior rumahdata uji ke-1 dikategorikan sebagai rumah tidak layak huni adalah sebagai berikut:

$$\begin{aligned}
P(\text{tidak layak} | X) &= P(\text{tidak layak}) \times P(X | \text{tidak layak}) \\
&= 0,443324937 \times 0,001405066 = 0,000622901
\end{aligned}$$

dan dikategorikan sebagai rumah layak huni adalah sebagai berikut:

$$\begin{aligned}
P(\text{layak} | X) &= P(\text{layak}) \times P(X | \text{layak}) \\
&= 0,556675063 \times 0 = 0
\end{aligned}$$

Hasil yang diperoleh data uji ke-1 diklasifikasikan ke dalam kelas 1 atau tidak layak. Ketepatan hasil klasifikasi metode Naive Bayes ditampilkan seperti berikut:

**Tabel 11.** Ketepatan Hasil Klasifikasi Naive Bayes

Kelas Asli	Kelas Hasil Klasifikasi	
	Kelas 1	Kelas 2
Kelas 1	19	0
Kelas 2	2	21

Berdasarkan Tabel 11 dapat dihitung nilai akurasi dan laju error sebagai berikut:

$$\begin{aligned}
\text{Akurasi} &= \frac{\text{jumlah data yang diklasifikasikan secara benar}}{\text{jumlah semua data yang diklasifikasikan}} \\
&= \frac{f_{11}+f_{22}}{f_{11}+f_{12}+f_{21}+f_{22}} = \frac{19+21}{19+0+2+21} = \frac{40}{42} = 0,9524 \\
\text{Laju error} &= \frac{\text{jumlah data yang diklasifikasikan secara salah}}{\text{jumlah semua data yang diklasifikasikan}} \\
&= \frac{f_{12}+f_{21}}{f_{11}+f_{12}+f_{21}+f_{22}} = \frac{0+2}{19+0+2+21} = \frac{2}{42} = 0,0476
\end{aligned}$$

Ketepatan klasifikasi metode Naive Bayes sebesar 0,9524 atau 95,24% dengan laju error sebesar 0,0476 atau 4,76%.

#### 4.4. Perbandingan Ketepatan Klasifikasi

Hasil perbandingan ketepatan klasifikasi dari metode LVQ dan Naive Bayes dapat dilihat sebagai berikut:

**Tabel 12.** Perbandingan Ketepatan Klasifikasi

Latih : Uji	LVQ	Naive bayes
95% : 5%	71,43%	95,24%
90% : 10%	66,67%	94,05%
80% : 20%	65,27%	91,62%
70% : 30%	62,15%	90,04%

Metode Naive Bayes bekerja lebih baik dibandingkan dengan LVQ dalam mengklasifikasikan rumah layak huni.

## 5. KESIMPULAN

Berdasarkan hasil dan pembahasan diperoleh kesimpulan hasil klasifikasi rumah layak huni dengan menggunakan metode *Learning Vector Quantization* adalah 30 rumah tepat klasifikasi dan 12 rumah tidak tepat klasifikasi. Ketepatan klasifikasi terbaik pada metode *Learning Vector Quantization* dengan proporsi pembagian data 95% data latih dan 5% data uji sebesar 71,43% dan laju error sebesar 28,57%. Hasil klasifikasi rumah layak huni dengan menggunakan metode *Naive Bayes Classifier* adalah 40 rumah tepat klasifikasi dan 2 rumah tidak tepat klasifikasi. Ketepatan klasifikasi terbaik pada metode *Naive Bayes Classifier* dengan proporsi pembagian data 95% data latih dan 5% data uji sebesar 95,24% dan laju error sebesar 4,76%. Berdasarkan nilai akurasi yang diperoleh, metode *Naive Bayes Classifier* menghasilkan nilai akurasi lebih tinggi dibandingkan metode *Learning Vector Quantization*. Sehingga metode *Naive Bayes Classifier* lebih baik dalam mengklasifikasikan rumah layak huni di Kabupaten Brebes.

## DAFTAR PUSTAKA

- [1] [BPS] Badan Pusat Statistika. 2011. *Indikator Perumahan dan Kesehatan Lingkungan 2011*. Jakarta: BPS
- [2] [BPS Jawa Tengah] Badan Pusat Statistika Jawa Tengah. 2013. *Profil Tempat Tinggal Jawa Tengah 2013*. Jawa Tengah: BPS
- [3] [BPS Jawa Tengah] Badan Pusat Statistika Jawa Tengah. 2014. *Jawa Tengah dalam Angka 2014*. Jawa Tengah: BPS
- [4] John, G. H. dan Langley, P. 1995. *Estimating Continuous Distributions in Bayesian Classifiers*. San Fransisco: Morgan Kaufmann Publishers Inc
- [5] Kusumadewi, S. 2004. *Membangun Jaringan Syaraf Tiruan (Menggunakan Matlab & Excel Link)*. Yogyakarta: Graha Ilmu
- [6] Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: ANDI
- [7] Prasetyo, E. 2014. *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: ANDI
- [8] Santoso, B. 2007. *DATA MINING: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu
- [9] Sujarweni, V. W. dan Endrayanto, P. 2012. *Statistika untuk Penelitian*. Yogyakarta: Graha Ilmu
- [10] Warsito, B. 2009. *Kapita Selektta Statistika Neural Network*. Semarang: BP Undip