

Perbandingan Metode SOM/Kohonen dengan ART 2 pada Data Mining Perusahaan Retail

Anak Agung Gede Bagus Ariana¹, I Ketut Gede Darma Putra², Linawati³

Abstract— This study investigates the performance of artificial neural network method on clustering method. Using UD. Fenny's customer profile in year 2009 data set with the Recency, Frequency and Monetary model data. Clustering methods were compared in this study is between the Self Organizing Map and Adaptive Resonance Theory 2. The performance evaluation method validation is measured by the index cluster validation. Validation index clusters are used, among others, Davies-Bouldin index, index and index Dunn Silhouette. The test results show the method Self Organizing Map is better to process the data clustering.

Index term— Data Mining, Artificial Neural Network, Self Organizing Map, Adaptive Resonance Theory 2.

Intisari—Penelitian ini ingin mengetahui unjuk kerja metode clustering data berbasis jaringan saraf tiruan. Menggunakan data set profil pelanggan UD. Fenny tahun 2009 dengan atribut Recency, Frequency dan Monetary. Metode clustering yang dibandingkan pada penelitian ini adalah Self Organizing Map dan Adaptive Resonance Theory 2. Evaluasi kinerja metode dilakukan dengan mengukur validasi index dari cluster yang terbentuk. Validasi cluster yang digunakan antara lain Indeks Davies-Bouldin, Indeks Dunn dan Indeks Silhouette. Hasil pengujian menunjukkan metode Self Organizing Map lebih baik dalam melakukan proses clustering data.

Kata Kunci— Data Mining, Jaringan Saraf Tiruan Self Organizing Map, Adaptive Resonance Theory 2.

I. PENDAHULUAN

Perusahaan ritel adalah perusahaan yang bergerak dalam bidang transaksi jual beli dalam jumlah kecil, satuan, atau eceran. Usaha ritel ini memiliki peranan penting dalam dunia perekonomian, karena tanpa usaha ritel suatu barang dari produsen tidak akan sampai di tangan konsumen. UD. Fenny merupakan perusahaan ritel yang menjual bahan kue, berlokasi di Jalan Nakula, Denpasar. Sejak tahun 2006 UD. Fenny telah menggunakan sistem informasi untuk menangani *point of sales* (POS) dan sistem *inventory*. Jumlah *record* transaksi ini bertambah secara signifikan dari tahun ke tahun. Jumlah *record* transaksi yang besar ternyata menyebabkan pihak manajemen sulit untuk menentukan keputusan-keputusan bisnis yang terkait dengan masalah strategi

pemasaran. Akibatnya perusahaan sering kehilangan pelanggan yang potensial menjadi pelanggan loyal.

Untuk mengatasi permasalahan tersebut, dilakukan proses *customer identification* atau pengenalan pelanggan. Salah satu cara yang dilakukan adalah *customer segmentation*. Pelanggan dapat disegmentasi dengan mengacu pada beberapa atribut. Salah satu model atribut yang sering digunakan adalah model RFM (*Recency*, *Frequency*, dan *Monetary*). *Recency* merupakan waktu pembelian akhir dari pelanggan yang diukur dalam satuan hari, bulan maupun tahun. *Frequency* menggambarkan jumlah total jumlah transaksi pelanggan yang diukur dalam rentang waktu tertentu, dan *monetary* menggambarkan jumlah nilai yang dibelanjakan oleh pelanggan [1] [2] [3].

Penelitian Nugraheni [4] telah membandingkan dua metode Fuzzy yang digunakan untuk proses *clustering* data yaitu metode Fuzzy C-Means dan metode Fuzzy Subtractive. Pemodelan data menggunakan model RFM. Evaluasi kinerja pembandingan menggunakan metode Modified Partition Coefficient dan Partition Entropy. Hasil perbandingan menunjukkan metode Fuzzy C-Means menghasilkan *cluster* yang lebih baik daripada metode Fuzzy Subtractive.

Menurut Smita [5] pada penelitian mengenai potensi penggunaan jaringan saraf tiruan pada data mining, menyebutkan metode SOM/Kohonen, Neuro Fuzzy dan Adaptive Resonance Theory 2 sebagai metode yang memiliki kehandalan yang baik dalam melakukan proses *clustering*, namun belum dijelaskan secara spesifik perbandingan unjuk kerja ketiga metode tersebut.

Penelitian sebelumnya [6] oleh penulis telah berhasil membangun aplikasi data mining dengan menggunakan metode SOM/Kohonen dengan menggunakan data set yang sama. Pada penelitian ini belum dilakukan pembahasan mengenai unjuk kerja metode SOM/Kohonen. Model data yang digunakan pada penelitian ini adalah model RFM. Penelitian ini menekankan pada proses pembentukan data set, melalui tahapan pembangunan data *warehouse* yang siap untuk penelitian selanjutnya.

Penelitian sebelumnya [7] oleh penulis telah menerapkan penggunaan metode ART 2 untuk segmentasi pelanggan PDAM kabupaten Gianyar dengan menggunakan atribut data pemakaian air, golongan tarif PDAM dan diameter pipa. Pengukuran kualitas cluster menggunakan Indeks Validasi Silhouette. Mencapai nilai tertinggi mendekati 1 yaitu 0.95, hal ini menunjukkan metode ART 2 baik untuk melakukan clustering data pelanggan PDAM.

Berdasarkan pemaparan diatas, dilakukan perbandingan kinerja metode SOM/Kohonen dan ART 2 dalam melakukan *clustering* data. Penilaian tingkat unjuk kerja metode dalam menentukan *cluster* terbaik diukur dengan menggunakan validasi *cluster* Indeks Davies-Bouldin, Indeks Dunn, dan

¹Mahasiswa, Magister Teknik Elektro Universitas Udayana Gedung Pascasarjana Universitas Udayana Jl. PB Sudirman Denpasar-Bali 80232 (telp/fax: 0361-239599; email: gungariana86@gmail.com)

^{2, 3}Dosen, Magister Teknik Elektro Universitas Udayana, Gedung Pascasarjana Jl. PB Sudirman Denpasar-Bali 80232 (telp/fax: 0361-239599; e-mail: duglaire@yahoo.com)



Indeks Silhouette. Metode akan diuji dengan menggunakan data set profil pelanggan UD. Fenny tahun 2009.

II. SELF ORGANIZING MAP/KOHONEN

Jaringan saraf tiruan SOM/Kohonen merupakan salah satu model jaringan saraf tiruan yang menggunakan metode *unsupervised learning*. Artinya, jaringan tersebut melakukan pembelajaran tanpa bimbingan data *input-target* [8].

Jaringan ini terdiri dari dua lapisan (*layer*), yaitu lapisan *input* dan lapisan *output*. Setiap neuron dalam lapisan *input* terhubung dengan setiap neuron pada lapisan *output*. Setiap neuron dalam lapisan *output* merepresentasikan *cluster* dari *input* yang diberikan. Adapun algoritma metode SOM/Kohonen adalah sebagai berikut:

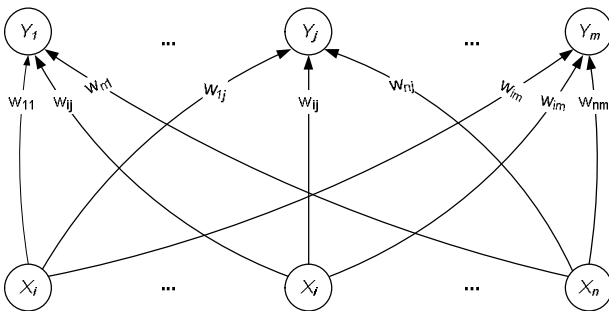
- 0. Neuron pada lapisan *input* (neuron *input*) sebanyak n dinotasikan sebagai x_1, x_2, \dots, x_n dan neuron pada lapisan *output* (neuron *output*) sebanyak m dinotasikan sebagai y_1, y_2, \dots, y_m . Bobot koneksi antara neuron *input* dan *output* dinotasikan sebagai w_{ij} ditentukan secara acak antara 0 dan 1.

- 1. Selama kondisi penghentian bernilai salah, lakukan langkah 2-7
- 2. Untuk setiap masukan x , lakukan langkah 3-5
- 3. Hitung semua j dengan menggunakan persamaan (1)

$$d_j = \sum_i (w_{ji} - x_i)^2 \tag{1}$$

- 4. Tentukan indeks j sedemikian hingga $d_{(j)}$ minimum
- 5. Untuk setiap unit j di sekitar j modifikasi bobot dengan persamaan (2)
- 6. Modifikasi laju pemahaman
- 7. Uji kondisi penghentian dilakukan dengan menghitung selisih antara bobot baru dengan bobot lama, apabila nilai berubah tidak signifikan, berarti pengujian sudah mencapai konvergensi sehingga dapat dihentikan.

$$w_{ji}^{baru} = w_{ji}^{lama} + \alpha(x_i - w_{ji}^{lama}) \tag{2}$$



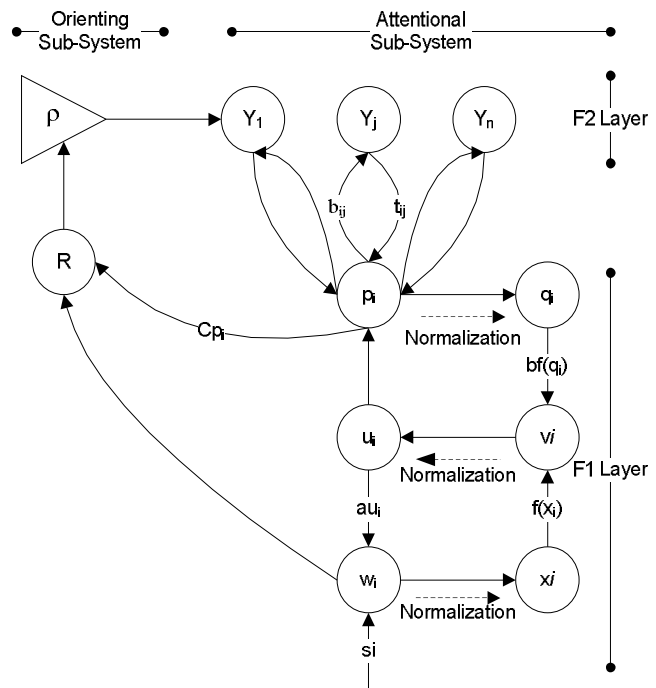
Gambar 1: Arsitektur Self-Organizing Map

III. ADAPTIVE RESONANCE THEORY 2

Adaptive Resonance Theory (ART) dirancang untuk mengatasi masalah kemampuan stabilitas-plastisitas (*stability-plasticity*) untuk menyimpan informasi yang telah dipelajari ketika ada informasi baru yang dihadapi oleh jaringan saraf tiruan lainnya [9] [10]. Arsitektur dasar dari ART 2 terdiri dari

Teknologi Elektro, Vol. 16, No. 02, Mei - Agustus 2017

tiga grup neuron yaitu: (1) *Input processing field - F1 layer* yang mengombinasikan sinyal-sinyal dari bagian masukan dan lapisan F_2 yang nantinya digunakan dalam membandingkan kemiripan sinyal masukan dengan vektor bobot untuk unit *cluster* yang telah dipilih sebagai calon yang diperbolehkan untuk mempelajari pola masukan yang dimasukkan. (2) *Cluster units - F2 layer* yang merupakan lapisan kompetitif, artinya unit *cluster* dengan jumlah masukan terbesar menjadi calon kuat yang diperbolehkan mempelajari pola masukan. (3) *Reset mechanism*. Suatu unit *cluster* diperbolehkan mempelajari pola masukan tergantung kepada kemiripan vektor *top-down*-nya dengan vektor masukan. Hal ini diatur oleh *reset* berdasarkan sinyal-sinyal yang diterimanya dari bagian masukan $F_{1(a)}$ dan bagian antarmuka $F_{1(b)}$ pada lapisan F_1 . [11] [12] [13]



Gambar 2: Arsitektur Adaptive Resonance Theory 2

Dalam algoritma ART 2 pemilihan parameter harus dilakukan dengan baik karena merupakan komponen penting dalam menggunakan algoritma ART 2 baik tipe pembelajaran cepat atau tipe pembelajaran lambat. Parameter dalam algoritma ART 2 terdiri dari:

- n : Jumlah komponen inputan.
- a, b : Nilai tetap atau konstanta.
- c : Diinisialisasi sebagai 0,1. Pemilihan nilai yang lebih kecil memberikan keefektifan terhadap nilai parameter *vigilance*.
- d : Aktivasi dari unit pemenang dalam layer F_2 , nilainya harus dipilih hampir mendekati 1 misalkan 0,9 karena akan memberikan keefektifan terhadap nilai parameter *vigilance*.

θ : *Noise suppression*, nilai ini didapat menggunakan persamaan (3) dan fungsi ini berhubungan dengan fungsi aktivasi (4).

$$\theta = 1/\sqrt{n} \quad (3)$$

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4)$$

α : *Learning rate*, atau laju pembelajaran.

ρ : *Vigilance factor* atau derajat kewaspadaan. Nilai parameter ini dapat diset dari 0 sampai 1 namun pemberian nilai parameter kurang dari 0,7 adalah kurang efektif karena sama saja artinya dengan nilai 0

b_j : Nilai bobot *bottom-up* dapat ditentukan dengan menggunakan persamaan (5).

$$b_j \leq \frac{1}{(1-d)\sqrt{n}} \quad (5)$$

t_j : Inisialisasi nilai bobot *top-down* menggunakan nilai yang kecil sehingga dapat mencegah reset terjadi pada input pertama yang masuk dalam suatu *cluster*.

Berikut adalah algoritma pembelajaran ART 2:

0. Inisialisasi parameter $a, b, \theta, c, d, e, \alpha, \rho$.

1. Lakukan langkah 2-12 untuk setiap *epoch* training.

2. Untuk setiap masukan vektor s , lakukan langkah 3-11

3. Modifikasi unit aktivasi F_1

$$u_i = 0 \quad (6)$$

$$q_i = 0 \quad (7)$$

$$p_i = 0 \quad (8)$$

$$x_i = \frac{s_i}{e + \|s\|} \quad (9)$$

$$w_i = s_i \quad (10)$$

$$v_i = f(x_i) \quad (11)$$

Modifikasi unit aktivasi F_1 lagi.

$$u_i = \frac{v_i}{e + \|v\|} \quad (12)$$

$$w_i = s_i + \alpha u_i \quad (13)$$

$$p_i = u_i \quad (14)$$

$$x_i = \frac{w_i}{e + \|w\|} \quad (15)$$

$$q_i = \frac{p_i}{e + \|p\|} \quad (16)$$

$$v = f(x_i) + bf(q_i) \quad (17)$$

4. Hitung sinyal unit F_2

$$y_j = \sum_i b_{ij} p_i \quad (18)$$

5. Selama kondisi *reset = true* lakukan langkah 6-7

6. Untuk unit F_2 pilih Y_j dengan nilai paling besar.

7. Cek nilai *reset*

$$u_i = \frac{v_i}{e + \|v\|} \quad (19)$$

$$p_i = u_i + dt_{ji} \quad (20)$$

$$r_i = \frac{u_i + cp_i}{e + \|u\| + c\|p\|} \quad (21)$$

Jika $\|r\| < p - e$, maka

$$Y_j = -1 \text{ (Inhibit J)} \quad (22)$$

Nilai *reset = true*, ke langkah 5

Jika $\|r\| > p - e$, maka

$$w_i = s_i + \alpha u_i \quad (23)$$

$$x_i = \frac{w_i}{e + \|w\|} \quad (24)$$

$$q_i = \frac{p_i}{e + \|p\|} \quad (25)$$

$$v_i = f(x_i) + bf(q_i) \quad (26)$$

Nilai *reset = false*, ke langkah 8

8. Lakukan langkah 9-11 untuk jumlah iterasi

9. Modifikasi bobot untuk pemenang unit J .

$$t_{ji} = \alpha du_i + \{1 + \alpha d(d-1)\} t_{ji} \quad (27)$$

$$b_{ji} = \alpha du_i + \{1 + \alpha d(d-1)\} b_{ji} \quad (28)$$

10. Modifikasi aktivasi F_1

$$u_i = \frac{v_i}{e + \|v\|} \quad (30)$$

$$w_i = s_i + \alpha u_i \quad (31)$$

$$p_i = u_i + dt_{ji} \quad (32)$$

$$x_i = \frac{w_i}{e + \|w\|} \quad (33)$$

$$q_i = \frac{p_i}{e + \|p\|} \quad (34)$$

$$v_i = f(x_i) + bf(q_i) \quad (35)$$

11. Tes kondisi berhenti untuk tiap perubahan bobot

12. Tes kondisi berhenti untuk jumlah *epoch* yang telah ditentukan.

Fungsi aktivasi yang digunakan adalah

$$f(x) = \begin{cases} x, & \text{if } x \geq \theta \\ 0, & \text{if } x < \theta \end{cases} \quad (36)$$

IV. METODE

Sebelum kinerja metode diukur, dilakukan pendefinisian tujuan pengujian, skenario pengujian dan parameter yang dibutuhkan dalam pengujian ini. Hasil pengujian dianalisis untuk menentukan pengaruh-pengaruh parameter pada masing-masing metode dan menentukan metode yang memiliki kinerja pengelompokan data yang lebih baik.

Pengujian yang dilakukan memiliki beberapa tujuan sebagai berikut:

1. Menganalisis pengaruh parameter *learning rate*, jumlah *cluster*, dan nilai *neighborhood* pada hasil *cluster* metode SOM.
2. Menganalisis pengaruh parameter *vigilance factor*, *noise suppression*, dan bobot pada hasil *cluster* metode ART 2.
3. Menganalisis tingkat kesempurnaan hasil masing-masing metode berdasarkan indeks validasi Davies-Bouldin Index, Dunn Index, dan Silhouette Index.

Skenario pengujian dilakukan agar proses pengujian dapat mencapai tujuan yang diinginkan. Skenario pengujian yang dilakukan yaitu dengan melakukan proses *clustering* pada *database* yang sudah dipersiapkan. Untuk mengetahui



pengaruh masing-masing parameter, maka dilakukan pengujian terhadap kombinasi nilai parameter pada masing-masing metode. Adapun parameter yang akan diberi nilai kombinasi pada metode SOM dan ART 2 adalah sebagai berikut:

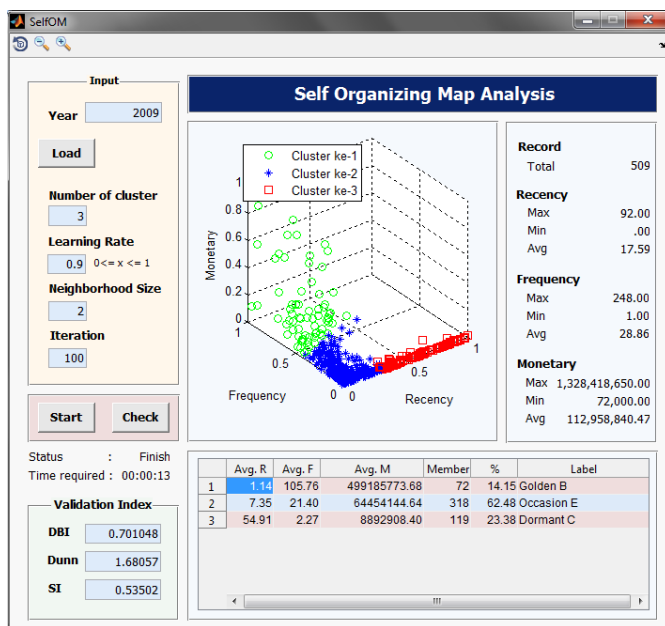
TABEL I
PARAMETER PENGUJIAN

| Metode | Parameter | Simbol | Kombinasi Nilai |
|--------|--------------------|----------|-------------------------|
| SOM | Jumlah cluster | nC | 2, 3, 4, 5 |
| | Learning rate | α | 0,2; 0,4; 0,6; 0,8; 0,9 |
| | Neighborhood value | Ne | 1, 2, 3, 4, 5 |
| ART 2 | Vigilance factor | ρ | 0,7; 0,8; 0,9; 0,95 |
| | Noise supression | θ | 0,1; 0,2; 0,3; 0,4; 0,5 |
| | Bobot | w | 1, 2, 3, 4, 5 |

Hasil cluster akan diuji dengan indeks validasi Davies-Bouldin Index, Dunn Index, dan Silhouette Index. Parameter pembentuk cluster terbaik berdasarkan pengujian masing-masing indeks validasi akan disimulasikan pada proses pengenalan kelas masing-masing cluster.

V. HASIL DAN PEMBAHASAN

Berikut adalah proses segmentasi dengan menggunakan metode SOM dengan parameter jumlah cluster = 3, learning rate = 0,9, dan neighbor size = 2.



Gambar 3: Antar Muka Metode SOM/Kohonen

Berikut adalah ringkasan nilai indeks validasi cluster terbaik berdasarkan kombinasi nilai parameter pada metode SOM/Kohonen.

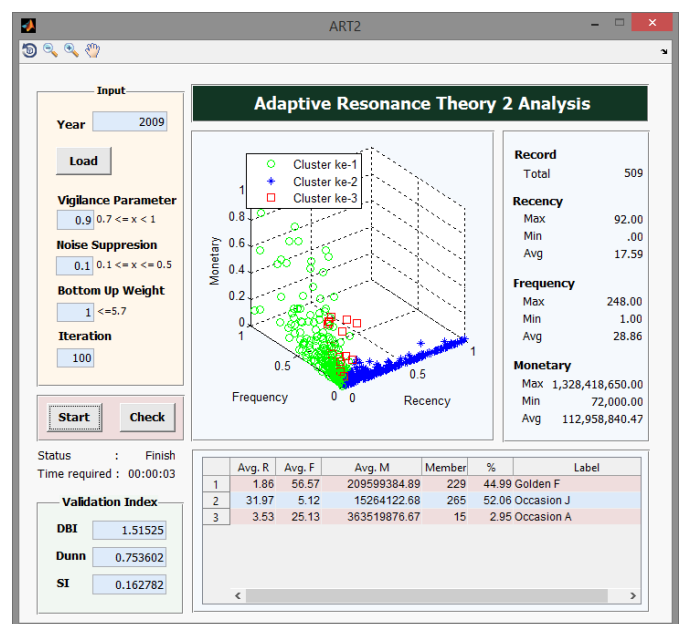
TABEL II
HASIL VALIDASI INDEX DATASET 2009 METODE SOM

| nC | α | Ne | Validasi | Nilai |
|----|-------------|---------|------------------|----------|
| 4 | 0,2; 0,8 | 1, 2 | Davies-Bouldin | 0,572172 |
| 2 | 0,2 s/d 0,9 | 1 s/d 5 | Dunn Index | 2,457409 |
| 3 | 0,2 s/d 0,9 | 1 s/d 5 | Silhouette Index | 0,53502 |

Berdasarkan keseluruhan percobaan dapat disimpulkan sebagai berikut:

1. Nilai indeks validasi cluster sangat dipengaruhi oleh jumlah cluster.
2. Nilai neighbor satu dan dua menghasilkan nilai indeks validasi yang lebih baik.
3. Nilai learning rate tidak terlalu mempengaruhi nilai indeks validasi.

Berikut adalah proses segmentasi dengan menggunakan metode ART 2 dengan parameter vigilance factor = 0,9; noise suppression = 0,1; dan bottom up weight = 1.



Gambar 4: Antar Muka Metode ART 2

Berikut adalah ringkasan nilai indeks validasi cluster terbaik berdasarkan kombinasi nilai parameter.

TABEL III
HASIL VALIDASI INDEX DATASET 2009 METODE ART 2

| nC | ρ | θ | w | Validasi | Nilai |
|----|----------|----------|---|----------------------|--------|
| 2 | 0,7; 0,8 | 0,1 | 1 | Davies-Bouldin Index | 1,0893 |
| 2 | 0,7; 0,8 | 0,1 | 1 | Dunn Index | 1,7959 |
| 2 | 0,7; 0,8 | 0,1 | 1 | Silhouette Index | 0,3307 |

Berdasarkan keseluruhan percobaan dapat disimpulkan sebagai berikut:

1. Nilai *vigilance factor* atau derajat kewaspadaan sangat mempengaruhi jumlah *cluster* yang terbentuk. Semakin tinggi nilai *vigilance factor* maka semakin banyak *cluster* yang dihasilkan.
2. Nilai bobot juga mempengaruhi jumlah *cluster* yang dihasilkan. Semakin besar nilai bobot maka *cluster* yang terbentuk akan lebih banyak.
3. Nilai *noise* berpengaruh terhadap rata-rata nilai validasi *cluster*. Nilai *noise* yang kecil menimbulkan nilai validasi *cluster* yang lebih baik.

Nilai indeks validasi menunjukkan tingkat kesempurnaan *cluster* yang terbentuk. Berikut adalah ringkasan nilai validasi indeks.

TABEL IV
PERBANDINGAN NILAI VALIDASI METODE SOM DAN ART 2

| Indeks Validasi | Unjuk Kerja | SOM | ART 2 |
|-----------------|-----------------------------|-----------------|--------|
| Davies-Bouldin | Lebih rendah semakin baik | 0,572172 | 1,0893 |
| Dunn | Lebih tinggi semakin baik | 2,457409 | 1,7959 |
| Silhouette | Mendekati satu semakin baik | 0,53502 | 0,3307 |

Tabel IV menunjukkan unjuk kerja metode SOM dinyatakan lebih baik daripada metode ART 2 diukur oleh validasi indeks Davies-Bouldin, indeks Dunn dan indeks Silhouette.

VI. KESIMPULAN

Berdasarkan hasil penelitian, nilai Davies-Bouldin Index terendah, Dunn Index tertinggi dan Silhouette Index yang mendekati satu didapatkan dari metode SOM/Kohonen. Jadi Metode SOM/Kohonen menghasilkan pengelompokan *cluster* dengan lebih baik.

REFERENSI

- [1] Tsipstsis, K., Chorionopoulos, A. (2009) *Data Mining Techniques in CRM*. United Kingdom: John Wiley and Sons
- [2] Cheng, C. H. & Chen Y. S. (2009) *Classifying the segmentation of customer value via RFM model and RS Theory. Expert System with Application*, 36, 4176-4184.
- [3] Hughes, A. M. (1994) *Strategic Database Marketing*. Chicago: Probus Publishing Company.
- [4] Nugraheni, Yohana (2013) Data Mining Using Fuzzy Method for Customer Relationship Management in Retail Industry. *Lontar Komputer* Vol. 4 No. 1, April 2013. ISSN: 2088-1541. Universitas Udayana.
- [5] Smita, N., (2010) *Potential use of Artificial Neural Network in Data Mining. Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference*. Vol.2, 339-343.
- [6] Ariana, A.A. Gede Bagus (2011) Customer Segmentation dengan Metode Self Organizing Map (Studi Kasus: UD. Fenny) *Lontar Komputer* Vol. 2 No. 1 Juni 2011. ISSN: 2088-1541. Universitas Udayana
- [7] Ariana, A.A. Gede Bagus., Dita Andriawan, I Wayan (2014) Segmentasi Pelanggan PDAM dengan Metode Adaptive Resonance Theory 2. *Prosiding Seminar Nasional Ilmu Komputer 2014*. ISBN: 978-602-19406-2-4. Universitas Gadjah Mada Yogyakarta.
- [8] Kohonen, T. (1990) The Self-Organizing Map, Invited Paper. *Proceedings of the IEEE*, Vol. 78, No. 9, September 1990.

- [9] Carpenter, G.A., & S. Grossberg (1987) ART 2: Self-organization of Stable Category Recognition Codes for Analog Input Patterns. *Applied Optics*. 26:4919-4930. A reprint from *Applied Optics* volume 26, number 23, December 1987.
- [10] Yao, X. (1999) Evolving Artificial Neural Networks, *Proceedings of the IEEE*, 7(9):1423-1447, September 1999.
- [11] Sivanandam, S. N., Sumathi, S. (2006) *Introduction to Data Mining and its Applications*. Springer, Verlag Berlin Heidelberg.
- [12] Sivanandam, S. N., Sumathi, S., & Deepa, S. N. (2006) *Introduction to neural networks using MATLAB 6.0*. New Delhi: The McGraw-Hill Companies.
- [13] Fausset, Laurene (1994) *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*. Prentice-Hall, New Jersey, USA.

