

# KOMBINASI SELF-ORGANIZING MAP NEURAL NETWORK DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI OTOMATIS CITRA KELOMPOK BINTIK MATAHARI

Gregorius Satia Budhi<sup>1</sup>, Rudy Adipranata<sup>2</sup>, Bachtiar Anwar<sup>3</sup>, Bambang Setiahad<sup>4</sup>, Adrian Hartanto N.<sup>5</sup>, Alvin Nathaniel T.<sup>6</sup>  
<sup>1,2,5,6</sup> Universitas Kristen Petra Surabaya.

E-mail: greg@petra.ac.id, rudya@petra.ac.id

<sup>3</sup> Pusat Pemanfaatan Sains Antariksa, Lembaga Penerbangan dan Antariksa Nasional (LAPAN)

E-mail: bachtiara@yahoo.com

<sup>4</sup> Stasiun Pengamat Dirgantara (SPD) LAPAN

E-mail: bambangsetiahad@rocketmail.com

**Abstract:** Observation and forecasting of weather disturbances in the earth and space caused by the appearance of sunspots is one of the tasks of the LAPAN. The sunspots are actually a tremendous explosion (flares) and coronal mass ejection (CME) that may affect electromagnetic field in a location on the earth. In order to support an early warning system that will be made, LAPAN wishes to make a computer application that can automatically classify a group of sunspots. Here we classify groups of sunspots on a sun digital image based on the "Modified - Zurich Sunspot Classification System". In this paper we present our experiment results on the sunspots group classification using the combination of Self-Organizing Map Neural Networks and K-Nearest Neighbor. The result of testing is promising because the classification accuracy was 100% for the data that is trained and 89.1% for the data that is not trained.

**Keywords:** Modified - Zurich Sunspot Classification, Self-Organized Map, K-Nearest Neighbors

## 1. Pendahuluan

Matahari merupakan sumber energi tak terbatas bagi kehidupan di bumi. Namun, selain sebagai sumber energi, matahari juga memberikan 'gangguan' ke lingkungan antariksa sekitar bumi serta kehidupan di bumi dalam bentuk badai matahari (*solar storm*). Dengan semakin meningkatnya ketergantungan kehidupan manusia dengan teknologi modern, baik fasilitas di permukaan bumi dan antariksa di sekitar bumi, kemunculan 'gangguan' dari matahari harus diantisipasi.

Sumber gangguan dari matahari diantaranya berupa ledakan dahsyat (*flare*) dan pelontaran massa korona (*Coronal Mass Ejection / CME*). Kedua tipe gangguan ini pada umumnya berasal dari kelompok bintik matahari (*sunspot*). Bintik matahari berevolusi dari bintik kecil dengan aktivitas rendah berkembang menjadi konfigurasi yang sangat kompleks dengan kemungkinan mempunyai aktivitas tinggi, yaitu mengeluarkan ledakan dan pelontaran massa korona. Untuk mengetahui tingkat kompleksitas bintik matahari dan aktivitasnya digunakan klasifikasi bintik matahari *Modified - Zurich* guna mengklasifikasikan jenis bintik matahari tersebut. Informasi mengenai klasifikasi ini sangat penting untuk menganalisa cuaca antariksa (*space weather*).

Penelitian ini adalah bagian dari penelitian yang lebih besar dengan judul "Klasifikasi Otomatis Kelompok Bintik Matahari Untuk Menganalisa Kondisi Cuaca Antariksa" yang kami lakukan berkerjasama dengan peneliti-peneliti dari Lembaga Penerbangan dan Antariksa Nasional (LAPAN). Klasifikasi kelompok bintik matahari (*sunspot*) ini berdasarkan sistem klasifikasi "*Modified - Zurich Sunspot Classification*". Dalam makalah ini kami mempublikasikan sebagian hasil penelitian kami yaitu penggunaan kombinasi metode jaringan saraf tiruan *Self-Organizing Map* dan metode data mining *K-Nearest Neighbor* guna proses klasifikasi kelompok bintik matahari (*cluster of sunspot*) yang muncul di dalam citra digital bintik matahari. Citra digital tersebut diambil dengan instrumen *Michelson Doppler Image (MDI) Continuum/SOHO (Solar dan Helispheric Observatory)*. Sebelum dikenai proses klasifikasi, pada citra digital bintik matahari tersebut perlu dilakukan perbaikan citra, deteksi lokasi bintik matahari dan pengelompokkan bintik matahari. Semua proses tersebut telah kami publikasikan sebelumnya [1, 6].

## 2. Tinjauan Pustaka

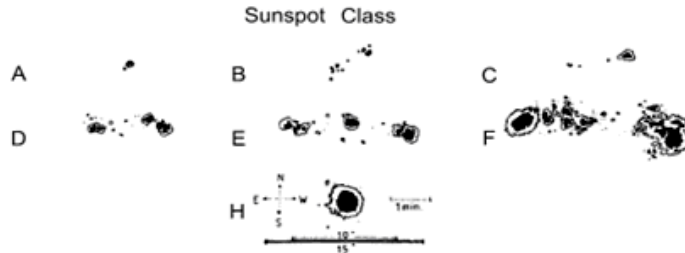
### 2.1 Klasifikasi Bintik Matahari (*Sunspot*)

Bintik matahari (*sunspot*) merupakan perpotongan tabung medan magnet matahari (*magnetic flux tube*) dengan fotosfer. Bintik matahari tampak hitam karena medan magnet mempunyai efek pendinginan sehingga suhu bintik matahari lebih rendah dari sekitarnya. Medan magnet bintik matahari terbukti merupakan sumber energi gangguan dari matahari, misalnya ledakan dahsyat (*flare*) atau pelontaran massa korona (*Coronal Mass Ejection / CME*) [14]. Fenomena ini dapat mengganggu lingkungan antariksa di sekitar Bumi maupun fasilitas teknologi dan kehidupan di permukaan Bumi [3, 4, 9].

Karena bintik matahari merupakan perpotongan tabung medan magnet dengan fotosfer, pemunculan bintik matahari diawali dengan sebuah bintik hitam berukuran kecil. Seiring dengan waktu, bila tabung medan magnet terus-menerus keluar dari

dalam matahari akibat gaya apung (*buoyancy force*), maka akan tampak dua buah bintik dengan polaritas medan magnet berlawanan (*bipolar*). Bintik matahari berkembang menjadi konfigurasi lebih kompleks, yaitu jumlah bintik dan luas bertambah. Semakin kompleks suatu konfigurasi bintik matahari, semakin besar kemungkinan terjadi ketidakstabilan (*instability*) medan magnet sehingga memicu peristiwa *flare* atau CME [2, 13].

Tahapan evolusi bintik matahari dinyatakan dalam “*Modified - Zurich Sunspot Classification*”, yaitu kelas A, B, C, D, E, F (tingkat kompleksitas bertambah) dan kemudian berangsur-angsur menurun hingga akhirnya menjadi kelas H. Gambar klasifikasi terdapat pada Gambar 1 [11].



Gambar 1. Klasifikasi “*Modified - Zurich Sunspot Classification*” [11]

## 2.2 Self-Organizing Map

*Self-organized maps* adalah sebuah metode jaringan syaraf tiruan yang pertama kali dikembangkan oleh Profesor Teuvo Kohonen sehingga sering juga disebut dengan *Kohonen-SOM*. Langkah-langkah training self-organizing maps adalah sebagai berikut [7]:

1. Inisialisasi semua *weight*.
2. Tentukan parameter *neighbourhood*.
3. Tentukan parameter untuk *learning rate*.
4. Selama *stopping condition* belum terpenuhi, lakukan langkah 5 sampai 8.
5. Untuk setiap input vector  $x$ , lakukan langkah 6 sampai 8.
6. Untuk setiap  $j$ , hitung *distance* dengan perhitungan:

$$D(j) = \sum_i (w_{ij} - x_i)^2 \quad (1)$$

7. Cari index  $J$  dimana  $D(J)$  bernilai minimum.
8. Untuk setiap  $j$  dalam *neighborhood* dari  $J$  dan untuk setiap  $i$ , lakukan perhitungan:

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha[x_i - w_{ij}(\text{old})] \quad (2)$$

9. *Update learning rate* dari *self-organized maps*.
10. *Update radius neighborhood*.
11. Cek apakah *stopping condition* terpenuhi, bila *true*, lanjut ke langkah 12.
12. Simpan bobot akhir.

### 2.2.1 Modifikasi metode training SOM

Untuk menggantikan langkah ke-6 dari algoritma di atas, digunakan persamaan [15]:

$$m_i(t+1) = m_i(t) + \alpha(t)h_{bi}(t)[x - m_i(t)] \quad (3)$$

$$h_{bi}(t) = \exp\left(-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}\right) \quad (4)$$

dimana  $r_b$  dan  $r_i$  adalah posisi *neuron*  $b$  (*best-matching unit*) dan  $i$  pada *Self-Organized Map* yang dibuat, dan  $\sigma$  adalah radius dari *neighborhood*.

Untuk menghitung *learning rate*  $\alpha$  digunakan persamaan berikut [5]:

$$\alpha(t) = \alpha_0 \exp\left(-\frac{t}{\lambda}\right) \quad (5)$$

Untuk menghitung *neighborhood* radius digunakan berikut [5]:

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\lambda}\right) \quad (6)$$

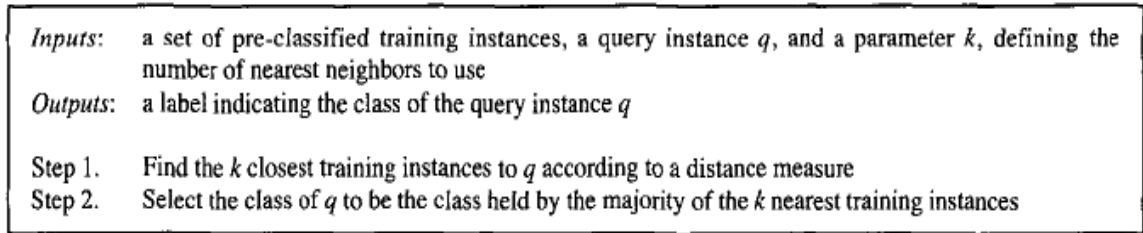
$$\sigma_0 = \frac{\max(\text{width}, \text{height})}{2} \quad (7)$$

$$\lambda = \frac{T}{\log(\sigma_0)} \quad (8)$$

dimana  $t$  adalah iterasi sekarang dan  $T$  adalah jumlah maksimal iterasi yang telah di setting sebelumnya.

### 2.3 k-Nearest Neighbors

Metode *k-Nearest Neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data yang jaraknya paling dekat dengan objek tersebut. Metode ini merupakan metode yang paling umum digunakan untuk estimasi dan prediksi [10]. Algoritma *k - Nearest Neighbors* dasar dapat dilihat pada Gambar 3.



Gambar 2. Algoritma *K-Nearest Neighbors* [12]

Untuk menghitung *distance measure* pada Step 1, digunakan rumus *eucledian distance*, pada Persamaan 1. Sementara itu untuk menghitung *majority vote* pada Step 2, digunakan rumus *weighted voting* seperti terlihat pada Persamaan 2 [10].

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (9)$$

$$vote = \frac{1}{d(x, y)^2} \quad (10)$$

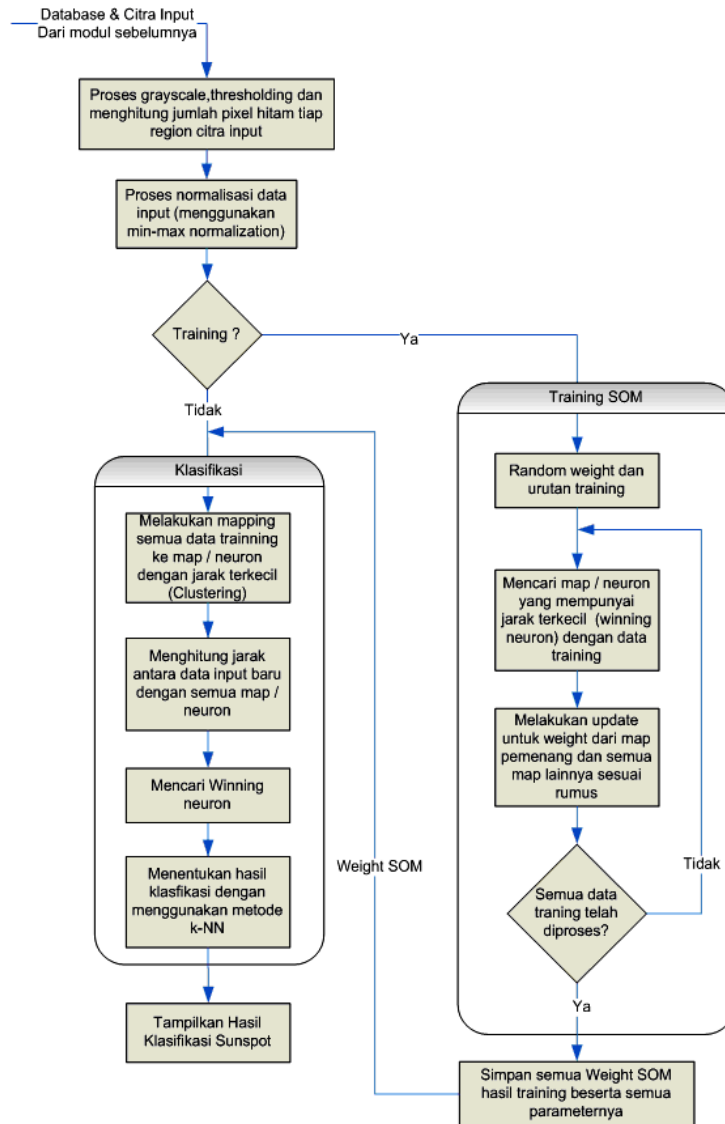
### 2.4 Min - Max Normalization

*Min-max normalization* adalah sebuah metode normalisasi yang mentransformasi data secara linear menjadi sebuah range yang baru [8]. Formula dari metode ini dapat dilihat pada Persamaan 11.

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A \quad (11)$$

## 3. Disain Sub-Sistem Klasifikasi Kelompok Bintang Matahari

Berikut ini adalah blok diagram dari disain prototipe yang dibangun untuk modul klasifikasi bintang matahari dengan metode jaringan saraf tiruan SOM.



Gambar 3. Blok Diagram dari sub-sistem klasifikasi bintik matahari dengan metode JST SOM.

Inputan dari sub-sistem ini berasal dari sub-sistem yang kita buat pada tahun pertama penelitian dan telah kita publikasikan [1, 6]. Setiap data yang masuk ke dalam sub-sistem klasifikasi bintik matahari ini terdiri dari 4 macam nilai dan satu citra digital. Empat macam nilai terdiri dari: data area (luas kelompok bintik matahari), perimeter (keliling kelompok bintik matahari), diameter kelompok bintik matahari, dan jumlah bintik yang ada pada satu kelompok bintik matahari. Sementara itu citra digital yang diinputkan adalah citra digital kelompok bintik matahari. Citra tersebut di-*crop* dari citra digital matahari oleh sub-sistem sebelumnya, contoh dapat dilihat pada Gambar 4. Selanjutnya citra digital kelompok bintik matahari itu dirubah menjadi data yang dapat diterima oleh neuron input dari JST SOM. Untuk merubah, citra ini diproses menjadi citra hitam putih dengan memanfaatkan metode *grayscaleing* dan *thresholding*. Setelah menjadi citra hitam-putih, dilakukan proses pembagian *region* pada citra, dimana jumlah *region* ditentukan oleh *user*. Contoh proses dapat dilihat pada Gambar 4. Selanjutnya jumlah *pixel* hitam pada tiap *region* akan dihitung.



Gambar 4. Contoh citra digital kelompok bintik matahari dan proses pembagian region

Berikutnya adalah proses normalisasi terhadap empat data input dan nilai - nilai yang didapat dari menghitung *pixel* hitam tiap *region*, menggunakan metode *Min - Max Normalization*. Nilai - nilai yang telah dinormalisasi inilah yang digunakan sebagai masukan bagi *neuron input* JST SOM.

Setelah semua data input telah siap, maka user dapat memilih proses yang akan dijalankan, yaitu: training JST atau klasifikasi. Untuk training JST, seperti pada umumnya proses training JST SOM, semua data sample akan di-training-kan sampai semua *weight* yang menghubungkan semua neuron input ke semua map SOM menjadi konvergen. Setelah konvergen, maka *weight* tersebut akan disimpan dan nantinya dapat digunakan pada proses klasifikasi. Untuk proses klasifikasi, inputan yang dibutuhkan adalah kumpulan *weight SOM* yang telah konvergen, arsitektur SOM yang dipakai, data - data *sample* yang digunakan mentraining SOM serta data - data kelompok bintang matahari yang akan diklasifikasi. Langkah pertama adalah melakukan *clustering* terhadap semua data sample sehingga akan terbentuk *cluster - cluster* data *sample* pada *map - map* SOM yang tersedia. Kemudian data yang akan diklasifikasi diproses pula. Disini akan dicari *winning map / neuron* yang paling sesuai dengan data yang diklasifikasi. Setelah didapat, maka semua *sample* data yang ter-*cluster* pada *winning neuron / map* akan diproses menggunakan metode k-NN guna mendapatkan nilai klasifikasi kelompok bintang matahari berdasarkan sistem klasifikasi “*Modified - Zurich Sunspot Classification*”, yaitu kelas A, B, C, D, E, F atau H.

#### 4. Pengujian Sub-Sistem Klasifikasi Kelompok Bintang Matahari

Pengujian dilakukan pada komputer dengan spesifikasi sebagai berikut:

Processor: Intel Pentium Quad Core i7

RAM: DDR3 4 GByte

O/S: Microsoft Windows 7

Database: Microsoft Access 2007

Total data sample: 214 kelompok bintang matahari

Dari hasil pengujian terhadap learning rate dapat ditarik kesimpulan bahwa learning rate saat training tidak terlalu berpengaruh pada hasil klasifikasi, asalkan weight yang dihasilkan telah konvergen. Sementara untuk pengujian banyak map SOM yang digunakan dapat ditarik kesimpulan bahwa map SOM 5 x 5 dan 7 x 7 cukup baik untuk proses clustering karena cluster - cluster data yang dihasilkan cukup mirip satu dengan lainnya. Sementara itu untuk hasil pengujian jumlah data sample dan pembagian region citra digital kelompok bintang matahari, terhadap kecepatan proses training dan akurasi klasifikasi dapat dilihat pada Tabel 1.

Tabel 1. Hasil pengujian Jumlah data training dan pembagian region terhadap lama training dan akurasi

Jumlah Data Training	Ukuran Pembagian Region	Lama Training (dalam millisecond)	Akurasi Klasifikasi Terhadap Data Training (dalam %)	Akurasi Klasifikasi Terhadap Data yang tidak ditraining (dalam %)
50	3 x 3	2871	100	77.44
50	5 x 5	6505	100	76.22
50	7 x 7	12074	100	74.39
100	3 x 3	5897	100	79.82
100	5 x 5	15554	100	70.16
100	7 x 7	31559	100	74.56
150	3 x 3	9298	100	89.06
150	5 x 5	24118	100	73.17
150	7 x 7	70044	100	87.5
214	3 x 3	14336	100	-
214	5 x 5	42729	100	-
214	7 x 7	100013	100	-

Dari pengujian pada Tabel 1 dapat disimpulkan bahwa, banyaknya data sample dan juga jumlah pembagian region akan meningkatkan waktu Training. Sementara untuk efek pada akurasi saat klasifikasi adalah sebagai berikut: SOM akan sangat akurat untuk sample data yang ditrainingkan. Sementara itu untuk pengujian dengan sample yang tidak ditrainingkan, semakin banyak data training semakin akurat hasil klasifikasi. Namun untuk pembagian region citra, semakin banyak region tidak mengakibatkan meningkatnya akurasi klasifikasi. Terbukti pada hasil pengujian terlihat bahwa pembagian region 3 x 3 menghasilkan akurasi yang lebih baik dari pembagian region 5 x 5 dan 7 x 7.

#### 5. Kesimpulan

Prototipe sub-sistem klasifikasi kelompok bintang matahari menggunakan metode Jaringan Saraf Tiruan Self-Organizing Map telah berhasil dibuat dengan benar. Hal ini dapat dilihat dari cepatnya proses trainingnya dan tingginya tingkat akurasi sub-

sistem ini untuk proses klasifikasi kelompok bintik matahari berdasarkan sistem klasifikasi "Modified - Zurich Sunspot Classification". Namun apakah nantinya sub-sistem ini akan digunakan atau tidak dalam aplikasi "Sistem Klasifikasi Otomatis Kelompok Bintik Matahari", belum kami putuskan. Hasil pengujian prototipe sub-sistem ini harus dibandingkan dengan hasil pengujian dua macam prototipe sub-sistem serupa, namun menggunakan metode jaringan saraf tiruan berbeda, yaitu: metode *Feed-Forward Backpropagation* dan *Probabilistic Neural Network*. Dari hasil perbandingan tiga macam prototipe sub-sistem tersebut akan ditentukan sub-sistem mana yang akan digunakan.

## 6. Pernyataan

Penelitian ini didanai dari Penelitian Hibah Bersaing DP2M Ditjen Dikti Depdiknas (101/SP2H/PP/DP2M/3/2010) dengan judul "Klasifikasi Otomatis Kelompok Bintik Matahari Untuk Menganalisa Kondisi Cuaca Antariksa".

## 7. Daftar Pustaka

- [1] Adipranata, R., Budhi, G.S., Setiahadi, B. dan Anwar, B. (2010). Segmentasi Bintik Matahari Menggunakan Metode Watershed. *Konferensi Nasional Sistem & Informatika Bali*.
- [2] Anwar, B., Acton, L.W., Makita, M., Hudson, H. S., McClymont, A.N. and Tsuneta, S. (1993). Rapid Sunspot Motion During A Major Solar Flare. *Solar Physics 147*, pp. 287-303, Kluwer Academic Publisher.
- [3] Boteler, D.H. (2001). Space Weather Effects on Power Systems. *Space Weather. Geophysical Monograph 125*, pp. 347.
- [4] Bothmer, V. and Daglis, I.A. (2007). *Space Weather, Physics and Effects*, Springer-Praxis Publishing.
- [5] Buckland, M. Kohonen's self organizing feature maps. <http://www.ai-junkie.com/ann/som/som1.html>. diakses terakhir tanggal 31 Maret 2011.
- [6] Budhi, G.S., Adipranata, R., Sugiarto, M., Anwar, B., dan Setiahadi, B. (2011). Pengelompokan Sunspot Pada Citra Digital Matahari Menggunakan Metode Clustering DbScan. *Prosiding Seminar Nasional Aplikasi Teknologi Informasi 2011, Yogyakarta*.
- [7] Fausett, L. (1994). *Fundamentals of neural networks: Architectures, algorithms, and applications*. Prentice-Hall International., Inc.
- [8] Han, J. and Kamber, M. (2006). *Data Mining: Concept and Techniques 2<sup>nd</sup> Edition*. McGraw Hill, Inc.
- [9] Lanzerotti, L.J. (2001). Space Weather Effects on Technologies. *Space Weather. Geophysical Monograph 125*, pp. 11.
- [10] Larose, D.T., (2005). *Discovering Knowledge in Data : An Introduction to Data Mining*. John Wiley & Sons, Inc.
- [11] McIntosh, P.S. (1990). The Classification of Sunspot Groups. *Solar Physics 125* vol. 125 no.2, pp. 251-267.
- [12] Schenker, A., Bunke, H., Last, M. and Kandel, A. (2005). *Graph-Theoretic Techniques For Web Content Mining*. World Scientific.
- [13] Setiahadi, B. (2005). Problems of Equilibria and Instabilities on Solar Coronal Magnetic Fields and Its Evolution Towards Energetic Energy Liberation: Effect to Interplanetary Space. *Prosiding Seminar Nasional Matematika, FMIPA UNDIP, E1, pp 1*.
- [14] Setiahadi, B., Sakurai, T., Miyazaki, H., and Hiei, E. (2006). Research on Magnetohydrodynamic Transport Phenomena in Solar-Terrestrial Space at LAPAN Watukosek 2006. *Prosiding Seminar Antariksa Nasional III, Pusat Pemanfaatan Sains Antariksa, LAPAN, Bandung, pp. 17*.
- [15] Vesanto, J. and Alhoniemi, E. (2000). Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, 11(3), 586-600. IEEE.