

# PENDEKATAN SELF-ORGANIZING MAPS DALAM DATA MINING UNTUK CLUSTERING PERUBAHAN HARGA SAHAM

*Self-Organizing Maps Approach in Data Mining for Clustering the Change of Stock Price*

Riris Prasetyo<sup>1</sup> dan Subanar<sup>2</sup>

Program Studi Ilmu Komputer  
Program Pascasarjana Universitas Gadjah Mada

## ABSTRACT

Money market is hoped can collect society's moneys for building and increasing society's income. Society as the main investor on money market has to know and understand the analysis of stock infestation for knowing how much of it, so they can give the most optimal return. One of the approaches for evaluating stock infestation is technical analysis that used data or note of market, which is published. For examples stock cost, market volume, the index of consolidations stock or individual, and the other factor which have technical characters.

The purpose of this research are for making a system which based to SOM for knowing what day the prices of stock are highest or lowest based on the frequent of each day appear and for knowing which algorithm is the most objective. This research use undirected data mining method that is clustering. Self-Organizing Maps (SOM) with training algorithm sequential and batch are used for clustering with output as like as clustering visualization. The result of this research show that the highest of stock prices is on Friday and the lowest is on Wednesday and Thursday.

**Key words :** *Data mining, clustering, Self-Organizing Maps.*

---

<sup>1</sup> STMIK Teknokrat, Lampung

<sup>2</sup> Fakultas MIPA I Universitas Gadjah Mada, Yogyakarta

## PENGANTAR

Pasar modal merupakan tempat kegiatan perusahaan dalam rangka mencari dana untuk pembiayaan usaha. Secara makro dalam pembangunan nasional memang menghendaki tersedianya dana untuk pembangunan dalam jumlah yang memadai, oleh karena itu dengan pasar modal diharapkan dapat menghimpun dana masyarakat untuk pembangunan sekaligus meningkatkan pendapatan masyarakat (Sumantoro, 1990). Masyarakat sebagai investor utama pada pasar modal diharuskan untuk mengetahui dan memahami analisis investasi saham untuk menentukan berapa perkiraan harga saham yang wajar sehingga dapat memberikan *return* paling optimal.

Salah satu pendekatan untuk penilaian investasi saham adalah analisis teknikal, pendekatan teknik analisis ini menggunakan data atau catatan tentang pasar yang dipublikasikan, seperti: harga saham, volume perdagangan, indeks harga saham gabungan maupun individu, serta faktor-faktor lain yang bersifat teknis. Fokus analisis teknikal menekankan pada pergerakan perubahan harga saham di dalam pasar atau suatu saham tertentu untuk jangka waktu yang relatif singkat, oleh sebab itu analisis lebih ditekankan untuk meramal trend perubahan harga, ini mengasumsikan bahwa perilaku harga masa lalu bisa merefleksikan harga dimasa yang akan datang (Sunariyah, 1997).

Seiring dengan kebangkitan kembali jaringan syaraf tiruan diakhir 1980-an yang disebabkan oleh kebutuhan pemrosesan informasi seperti otak, kemajuan teknologi komputer, dan perkembangan ilmu syaraf yang semakin mengerti mekanisme otak telah menjadikan jaringan syaraf tiruan sebagai salah satu alat untuk data mining. *Self-Organizing Maps (SOM)* yang ditemukan Kohonen *et al.*, (1995) dapat digunakan untuk melakukan *clustering* data yang besar maupun kecil dengan keluaran berupa visualisasi dengan cepat, sehingga visualisasi *clustering* ini dapat dengan cepat untuk mengamati dan menganalisis perubahan harga saham pada pasar modal.

Vesanto (2002) telah mengembangkan metode SOM untuk visualisasi dan analisis *clustering* dan kerangka kerja untuk analisis data. Berdasarkan hal ini penulis mengerjakan penelitian untuk melakukan visualisasi *clustering* pada beberapa harga saham emiten yang terdaftar dalam Bursa Efek Jakarta.

Tujuan yang ingin dicapai dalam penelitian ini adalah: (1)

digunakan untuk menganalisis harga-harga saham tertinggi dan terendah berdasarkan frekuensi masing-masing pemunculan hari dengan menggunakan algoritma sekuensial dan *batch*; (2) menentukan algoritma yang paling objektif diantara sekuensial dan *batch*.

Agar pembahasan nantinya tidak meluas, penulis dalam penelitian ini membatasi pada hal-hal: (1) parameter-parameter SOM yang digunakan yaitu: cara inialisasi *map*, topologi *map* dan ukuran *map* pada saat inialisasi; algoritma *training*, jari-jari *neighborhood*, rata-rata *learning* dan lama *training* pada saat *training*; (2) *clustering* yang dimaksud dalam penelitian ini adalah unit-unit *map* hasil visualisasi vektor-vektor prototipe dari hasil *training* SOM; (3) data yang digunakan adalah data harga saham emiten di Bursa Efek Jakarta pada periode waktu 2 Januari 2002 - 30 Juni 2002; dan (4) alat penelitian yang digunakan adalah SOM *Toolbox 2.0*

## LANDASAN TEORI

### Data Mining

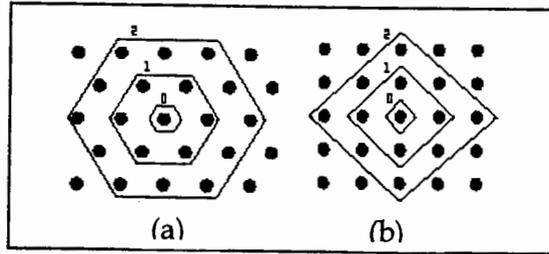
Data Mining adalah suatu proses eksplorasi dan analisis, dengan cara otomatis atau semi otomatis, dari sejumlah data yang besar supaya menemukan pola dan *rule* yang penuh arti (Berry dan Linoff, 2000).

Data mining digolongkan menjadi 2 bagian: (1) *directed data mining* menggunakan pendekatan atas-bawah, digunakan ketika tujuannya sudah diketahui, misalnya klasifikasi, estimasi dan prediksi; (2) *undirected data mining* menggunakan pendekatan bawah-atas yang mana membiarkan data untuk berbicara sendiri, tidak ada satu variabel sebagai target. Tujuannya adalah membentuk beberapa relasi diantara semua variabel, sehingga ditemukan suatu pola di dalam data dan menyerahkan ke pengguna untuk menentukan apakah pola tersebut penting atau tidak. Contohnya kelompok afinitas, *clustering* dan diskripsi atau visualisasi. Proses data mining meliputi *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment*.

### Algoritma SOM

Pada dasarnya SOM terdiri dari M unit lokasi pada *grid* berdimensi rendah yang beraturan, biasanya berdimensi 1 atau berdimensi 2. Lihat Gambar 1., *Grid-grid* berdimensi yang lebih tinggi adalah mungkin tetapi tidak umum digunakan selama visualisasinya

bermasalah. Tiap unit  $j$  memiliki suatu kumpulan vektor prototipe  $m_j = [m_{j1}, m_{jd}]$  yang berdimensi  $d$ . Posisi unit  $r_j$  pada *grid* ditetapkan dari awal. Pemilihan *map* terhadap data dengan menyesuaikan vektor prototipenya. Bersama *grid* dan himpunan bentuk vektor prototipe *map* yang berdimensi rendah yang macam datanya: suatu representasi 2-dimensi dimana sifat topologinya berhubungan erat dengan objek (unit-unit *map*) yang berdekatan satu sama lain (Vesanto, 2002).



Gambar 1. Dua SOM berukuran 5\*5 satuan. (a) *hexagonal lattice*, (b) *rectangular lattice* (dengan perbedaan nilai jari-jari *neighborhood*  $\sigma=0,1$  dan 2) (Sumber: Vesanto, 2000).

### Algoritma Training

Algoritma-algoritma *training* yang digunakan adalah

#### Sekuensial Algoritma Training,

Algoritma SOM adalah bersifat iterasi. Pada setiap langkah  $t$  *training* contoh vektor data  $x_i$  secara acak dipilih dari himpunan *training*. Jarak antara  $x_i$  dan seluruh vektor prototipe dihitung. *Best-matching unit (BMU)* dinyatakan dengan  $b_i$ , adalah unit *map* dengan prototipe paling dekat terhadap  $x_i$ .

$$b_i = \arg \min_j \{ \|x_i - m_j(t)\| \} \quad (2.1)$$

Jika contoh vektor  $x_i$  memiliki beberapa nilai yang hilang, variabel-variabel tersebut dilewati dalam perhitungan jarak dan dalam langkah *update subsequent*. Selanjutnya, vektor-vektor prototipe di *update* dengan memindahkannya menuju  $x_i$  seperti diperlihatkan pada gambar 2. Aturan *update* untuk vektor prototipe  $j$  adalah:

$$m_j(t+1) = m_j(t) + \alpha(t)h_{b_i,j}(t)[x_i - m_j(t)], \quad (2.2)$$

dimana  $t$  adalah indek langkah *training*,  $\alpha(t)$  adalah rata-rata *training* dan  $h_{b_i,j}(t)$  adalah pusat *neighborhood kernel* pada unit pemenang. *Kernel*

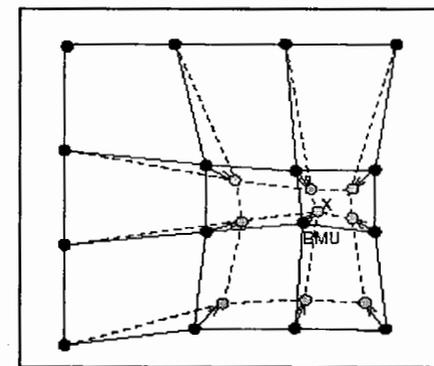
mendapat nilai terbesar untuk unit pemenang dan berkurang secara monoton dengan bertambahnya jarak pada *grid map*  $\|r_b - r_j\|$ . *Kernel* dapat dihitung misalnya dengan Gaussian:

$$h_{b_i,j}(t) = e^{-\frac{\|r_{b_i} - r_j\|^2}{2\sigma^2(t)}} \quad (2.3)$$

dimana  $r_{b_i}$  dan  $r_j$  adalah posisi unit  $b_i$  dan  $j$  pada *grid* SOM dan  $\sigma(t)$  adalah jari-jari *neighborhood*. Rata-rata *training*  $\alpha(t)$  dan jari-jari *neighborhood*  $\sigma(t)$  berkurang secara monoton selama *training*, rata-rata *training* menuju nol dan jari-jari *neighborhood* menuju ke beberapa nilai tak nol yang sesuai, seringkali bernilai satu.

Selama *training*, SOM bertindak seperti sebuah jaringan yang fleksibel yang berlipat-lipat menyerupai *cloud* yang dibentuk oleh data *training*. Karena hubungan *neighbor*, prototipe yang berdekatan ditarik ke arah yang sama, dan unit-unit yang berdekatan memperoleh vektor-vektor prototipe yang serupa.

Vektor-vektor prototipe menentukan suatu *tesselation* dari ruang *input* ke dalam suatu himpunan daerah Voronoi atau himpunan Voronoi  $V_j = \{x_i \mid \|x_i - m_j\| < \|x_i - m_k\| \forall k \neq j\}$  setiap hubungan terhadap satu unit *map*. Setiap efek vektor data di bawah himpunan Voronoi dari unit *map* ke arah paling dekat dengan vektor data (Vesanto dan Hollmen, 2002).



Gambar 2. *Updating best-matching unit (BMU)* dan *neighbor*nya menuju contoh *input* yang ditandai dengan  $x$ . Lingkaran hitam dan tidak hitam berhubungan dengan situasi sebelum dan sesudah *updating*. Garis tebal dan putus-putus menunjukkan hubungan *neighbor* (Sumber : Vesanto,2002).

## Batch Map

*Batch map* adalah suatu versi SOM yang berdasarkan pada iterasi tetap. Pada tiap langkah *training* BMU-BMU  $b_i$  dari seluruh vektor-vektor data diperoleh diawal dengan persamaan 2.1 Setelah ini prototipe baru dihitung dengan:

$$m_j = \frac{\sum_{i=1}^N h_{b_i,j} x_i}{\sum_{i=1}^N h_{b_i,j}} \quad (2.4)$$

Prototipe-prototipe baru rata-rata diberi bobot dari sampel data, sedemikian sehingga bobot tiap sampel data adalah nilai *kernel neighbor*  $h_{b_i,j}$  pada BMU  $b_i$  nya. Dasar SOM adalah suatu aproksimasi *stochastic* dari *batch map*.

Alternatifnya, *update* dapat dihitung sebagai bobot rata-rata

pusat himpunan Voronoi  $n_j = \frac{1}{N_j} \sum_{x_i \in V_j} X_i$ :

$$m_j = \frac{\sum_{i=1}^M N_i h_{ij} n_i}{\sum_{i=1}^M N_i h_{ij}} \quad (2.5)$$

dimana  $N_j$  adalah jumlah sampel-sampel dalam himpunan Voronoi  $N_j$ . Ini membolehkan suatu implementasi berbasis matrik yang lebih efisien dari pada persamaan 2.4 (Vesanto, 2002).

## Visualisasi Berbasis SOM

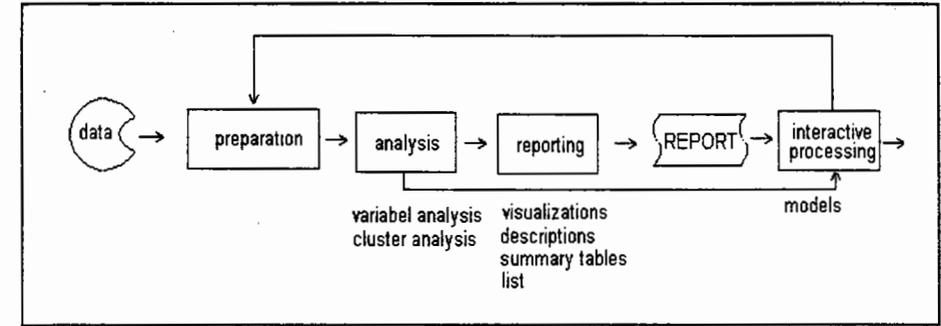
Visualisasi SOM menjalankan dua tugas yaitu: (1) dalam *cluster* dan teknik visualisasi prototipe : SOM dilihat sebagai representasi sampel dari data; dan (2) melihat SOM sebagai model dari data, dengan membandingkan data ke model ini, atau sampel data yang berbeda.

Visualisasi SOM berdasarkan pada tujuan visualisasi: (1) visualisasi bentuk dan struktur *cluster*; (2) visualisasi komponen atau variabel; dan (3) Visualisasi data pada *map*.

## Analisis format-tabel data

Proses eksplorasi data biasanya dikerjakan secara interaktif dan iteratif, sehingga ini memerlukan sejumlah pekerjaan secara manual. Terdapat 2 cara analisis tabel: (1) analisis sampel bertujuan

untuk mencari himpunan bagian dari bentuk data asli (atau minimal yang dapat memberi arti) dan karakterisasi sifat-sifat untuk memberikan *data miner* suatu ide dari struktur dan isi yang bermacam-macam. Dikerjakan dengan analisis *cluster*, karena analisis *cluster* menyediakan informasi yang bersifat kuantitatif dari struktur dan isi himpunan data; dan (2) analisis variabel dengan cara menggunakan



satu *component plane* untuk setiap variabel.

Gambar 3. Pemahaman data sebagai proses yang bersifat iterasi.

(Sumber : Vesanto (2000); Vesanto, 2002a)

## Clustering

Tujuan *clustering* adalah untuk membagi bermacam-macam group ke sejumlah subgroup yang memiliki kemiripan. Pada *clustering* tidak ada definisi kelas sebelumnya dan tidak ada contoh. *Clustering* sering dikerjakan mendahului beberapa bentuk data *mining* atau *modeling* (Berry dan Linoff, 2000).

## Saham

Menurut Kasmir (2002) saham adalah surat berharga yang bersifat kepemilikan. Pemain utama dalam pasar modal adalah perusahaan yang akan melakukan penjualan surat-surat berharga atau melakukan emisi dibursa biasanya disebut emiten dan pemodal yang akan membeli atau menanamkan modalnya diperusahaan yang melakukan emisi disebut investor.

## PROSES DATA MINING

### Data Preparation

Langkah-langkah *data preparation* dalam penelitian yaitu : (1) *select data*, dalam penelitian penulis menggunakan data harga saham

emiten pada Bursa Efek Jakarta pada periode 2 Januari 2002 – 30 Juni 2002; (2) *clean data*, membersihkan *error* dan memilih variabel-variabel yang menarik yaitu hari, tanggal, kode, harga penutupan saham; (3) *construct data*; (4) *integrate data*, dibuat beberapa tabel gabungan sesuai kelompoknya, adapun kelompok-kelompok tersebut adalah: kelompok saham perdagangan besar barang produksi, perdagangan eceran, jasa komputer dan perangkatnya, farmasi, tekstil dan garmen, plastik dan kemasan, pertambangan, makanan dan minuman, saham teraktif (Kompas, 2003), dan kelompok saham berdasarkan kepemilikan oleh asing kurang dari 49% dan yang lebih dari 49% (Siswanto, 2003); (5) *format data*, diasumsikan: banyaknya data sangat bervariasi, tetapi jumlah, tipe dan tujuan variabel dalam sampel telah diketahui sebelumnya. emiten sebagai variabel dan ditambah satu variabel hari aktif perdagangan saham (Senin - Jumat) sebagai label.

### Data Preprocessing

Dalam data *preprocessing* digunakan *linear scaling*, yaitu mentransformasikan skala jarak nilai variabel menjadi diantara [0 1] dengan rumus:  $x(\text{new}) = (x - \min(x)) / (\max(x) - \min(x))$ .

### Inisialisasi dan Training

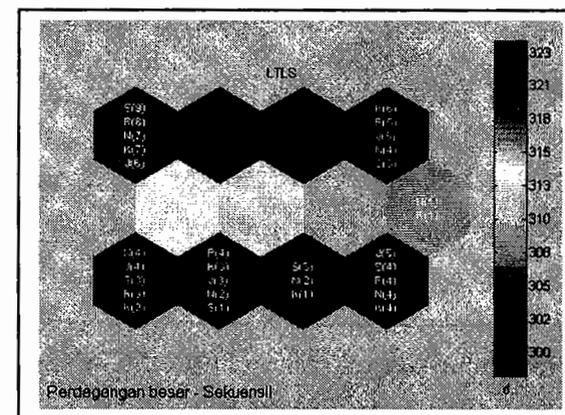
Paramater yang diperhatikan dalam menyusun *grid map* yaitu inisialisasi *map* topologi *map*, dan ukuran *map*. Selanjutnya menentukan algoritma *training* yaitu *sequential training* dan *batch training*. Parameter yang harus diperhatikan saat *training* yaitu jari-jari *neighborhood* ( $\sigma(t)$ ), rata-rata *learning* ( $\alpha(t)$ ), panjang *training*, dan fungsi *neighborhood*.

### Cara Interpretasi Hasil Visualisasi

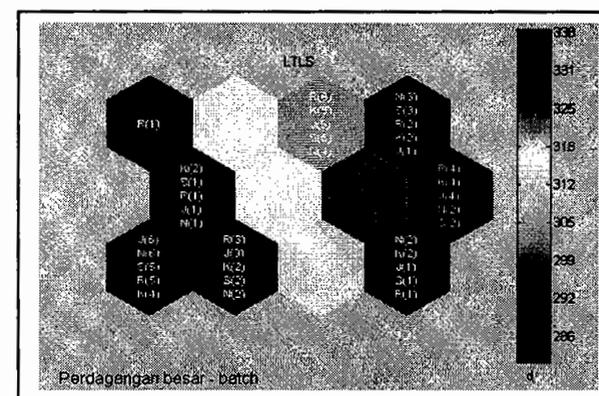
Untuk menginterpretasikan masing-masing gambar hasil visualisasi harga saham emiten yaitu dengan cara: (1) lihat degradasi warna yang tertinggi dan terendah berdasarkan *color bar*; (2) pada unit-unit *map* tersebut terdapat frekuensi hari pemunculan dalam kurun waktu 6 bulan. Hari-hari dalam unit *map* dikodekan dengan N: Senin, S: Selasa, R: Rabu, K: Kamis, dan J: Jumat. Sedangkan untuk frekuensi ditampilkan dengan angka yang ada dalam kurung setelah kode hari.

## HASIL DAN PEMBAHASAN

Berdasarkan pada tabel-tabel data yang telah dikelompokkan selanjutnya dengan menggunakan sistim SOM pada setiap tabel masing-masing *training* dengan algoritma sekuensial dan *batch* sehingga diperoleh visualisasi-visualisasi. Berikut ini diberikan salah satu contoh hasil visualisasi emiten Lautan Luas yang termasuk dalam kelompok perdagangan besar barang produksi.



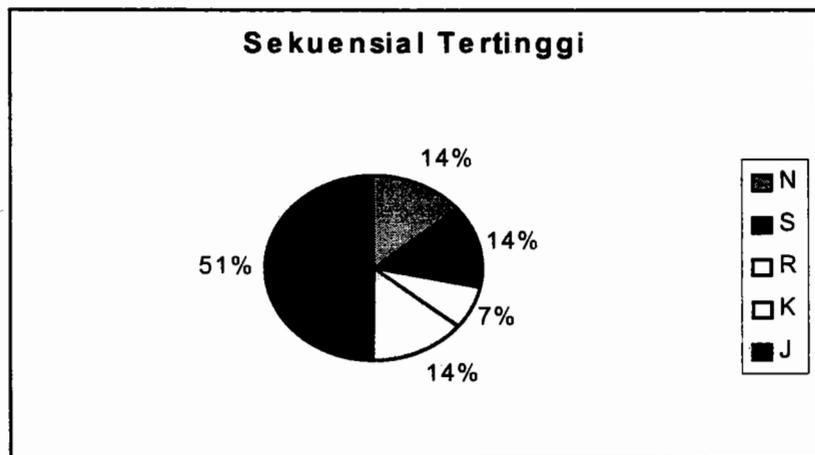
Gambar 4. Visualisasi emiten Lautan Luas dengan algoritma sekuensial



Gambar 5. Visualisasi emiten Lautan Luas dengan algoritma *batch*

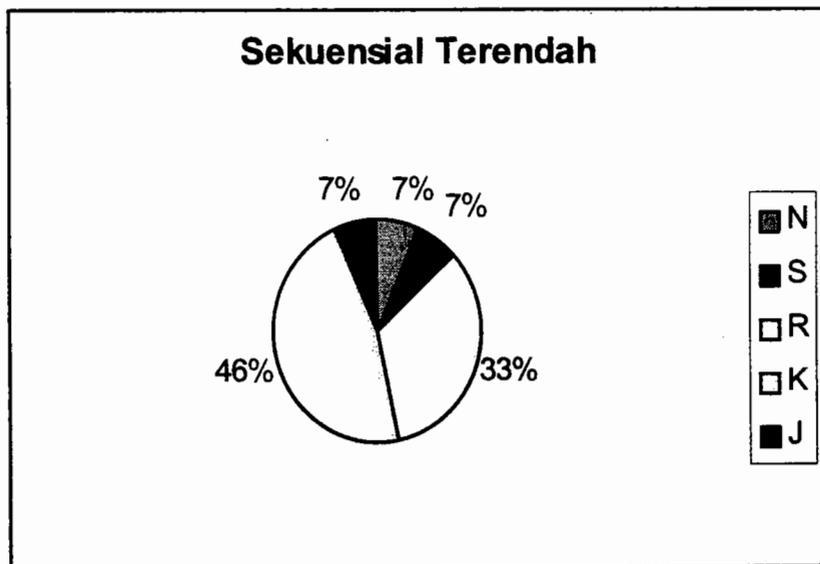
Berdasarkan Gambar 4. di atas diperoleh nilai saham tertinggi pada hari Senin dan Jumat dengan frekuensi pemunculan masing-

masing 4 kali. Sedangkan untuk nilai saham terendah terjadi pada hari Kamis dengan frekuensi pemunculan 6 kali. Begitupula untuk gambar 5 nilai saham tertinggi pada hari Senin dan Jumat dengan frekuensi pemunculan masing-masing 6 kali. Sedangkan untuk nilai saham terendah terjadi pada hari Rabu, Kamis, dan Jumat dengan frekuensi pemunculan masing-masing 4. Selanjutnya dari hasil keseluruhan

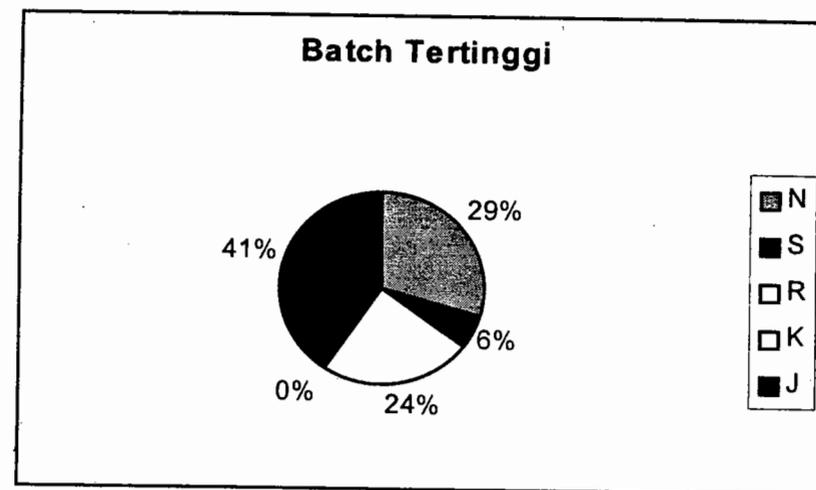


visualisasi maka diperoleh diagram seperti berikut (Gambar 6 s/d 9) :

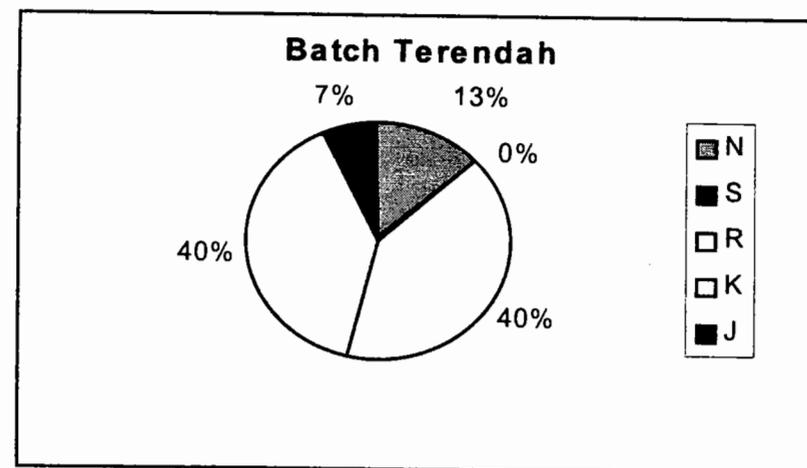
Gambar 6. Prosentase frekuensi hari tertinggi dengan algoritma sekuensial



Gambar 7. Prosentase frekuensi hari terendah dengan algoritma sekuensial



Gambar 8. Prosentase frekuensi hari tertinggi dengan algoritma batch



Gambar 9. Prosentase frekuensi hari terendah dengan algoritma batch

Berdasarkan gambar-gambar di atas maka dapat dikatakan baik untuk algoritma sekuensial maupun untuk algoritma batch frekuensi hari yang paling tinggi adalah Jumat. Frekuensi hari yang

paling rendah berdasarkan algoritma sekuensial adalah Kamis sedangkan algoritma *batch* adalah Rabu dan Kamis.

Selanjutnya berdasarkan hasil kedua visualisasi di atas dapat dikatakan bahwa algoritma yang paling objektif adalah algoritma *batch* hal ini dikarenakan dalam algoritma *batch* tidak terdapat nilai  $\alpha$  (rata-rata *learning*) yang diinputkan berdasarkan keinginan *user* yang berarti unsur subjektif tidak terlibat. Sedangkan untuk algoritma sekuensial berlaku sebaliknya. Penarikan kesimpulan bahwa algoritma yang paling objektif adalah *batch* semata-mata karena tahapan analisis data mining masih berada dalam tahap *explanatory*.

## KESIMPULAN

### Kesimpulan

Berdasarkan hasil pembahasan maka dapat disimpulkan:

1. Sistem data mining berbasis SOM yang dihasilkan digunakan untuk menganalisis data harga saham dari BEJ periode 2 Januari - 30 Juni 2002 untuk menentukan harga-harga tertinggi dan terendah berdasarkan frekuensi masing-masing pemunculan hari dengan hasil sebagai berikut.
  - Harga-harga saham tertinggi terjadi pada hari Jumat baik dengan menggunakan algoritma sekuensial maupun algoritma *batch*.
  - Harga-harga saham terendah dengan algoritma sekuensial terjadi pada hari Kamis dan dengan algoritma *batch* terjadi pada Rabu dan Kamis.
2. Algoritma yang paling objektif adalah *batch*.

### Saran

Setelah dilakukan penelitian maka peneliti memberikan saran sebagai berikut.

1. Agar ada penelitian lebih lanjut tentang algoritma mana yang paling efektif dan efisien diantara algoritma sekuensial dan *batch*.
2. Hendaknya ada penelitian lebih lanjut dengan pendekatan analisis teknikal dengan menggabungkan harga saham dengan data lainnya seperti volume perdagangan, indeks harga saham gabungan maupun individu serta faktor-faktor lain yang bersifat teknis.

## DAFTAR PUSTAKA

- Berry, Michael J. A dan Linoff, G. S., 2000, "Mastering Data Mining: The Art and Science of Customer Relationship Management", John Wiley & Sons, Inc
- Kasmir, 2002, "Bank dan Lembaga Keuangan lainnya", PT. Raja Grafindo Persada, Jakarta.
- Siswanto, Heri, 2003, "Dampak Kepemilikan Saham-saham Oleh Asing Terhadap Pengaruh Faktor International di PT. Bursa Efek Jakarta (Periode tahun 2002)", UGM, Yogyakarta
- Sumantoro, 1990, "Pengantar tentang Pasar Modal di Indonesia", Ghalia Indonesia, Jakarta, pp 9-61
- Sunariyah, 1997, "Pengantar Pengetahuan Pasar Modal", UPP AMP YKPN, Yogyakarta, pp 1-92
- Vesanto, J., 2002, "Data Exploration Process Based on the Self-Organizing Map", Helsinki University of Technology
- Vesanto, J., 2000, "Using SOM in Data Mining", Helsinki University of Technology
- Vesanto, J. dan Hollmén, J., 2002(a), "An Automated Report Generation Tool for the Data Understanding Phase", In Hybrid Information Systems, edited by A. Abraham and M. Köppen, Physica Verlag, Heidelberg, pp. 611-626. © 2002 Springer-Verlag.
- <http://lib.hut.fi/Diss/2002/isbn9512258978/index.html>