

# PERAMALAN MULTI ATRIBUT DENGAN MENGGUNAKAN FUZZY CLUSTERING (STUDI KASUS: STOCK PRICE)

M. ANDY FARISKA 7406.040.029

Dosen Pembimbing :

ARNA FARIZA, S.Kom, M. Kom

NIP. 197107081999032001

YULIANA SETYOWATI., S.Kom, M. Kom

NIP. 197807062002122003

Jurusan Teknik Informatika  
Politeknik Elektronika Negeri Surabaya  
Institut Teknologi Sepuluh Nopember,  
Surabaya

E-mail: 1) [vixionviper@gmail.com](mailto:vixionviper@gmail.com)

---

## ABSTRAK

Sektor ekonomi merupakan salah satu sektor yang paling dekat dengan manusia, yang dimana salah satu bentuk yang pada saat ini sedang digemari adalah pasar modal atau yang biasa disebut sebagai bursa saham. Bursa saham dapat dilakukan oleh individu maupun Institusi keuangan pada perusahaan yang sudah go public. Saham sendiri merupakan suatu aset yang mudah untuk diperjual belikan dibandingkan dengan aset konvensional seperti tabungan dll. Akan tetapi pergerakan nilai dari suatu saham adalah dinamis yang dimana dapat berubah sewaktu-waktu. Investasi jangka seperti saham dapat dianalisa secara akurat dengan menggunakan analisa teknikal dalam proyek akhir ini.

Dalam sebuah permasalahan, dapat dianalisa pola dari suatu permasalahan tersebut terhadap faktor yang mempengaruhinya dengan metode fuzzy. Permasalahan saham yang di bahas pada proyek akhir ini adalah sebuah permasalahan time series yang dimana dapat dianalisa dari variabel – variabel yang disusun berdasarkan mengikuti urutan waktu. Hal ini sangat penting karena suatu data history merupakan faktor yang sangat penting apabila tidak memahami faktor – faktor yang mempengaruhi sistem. Solusi yang diusahakan sudah berbagai macam mulai dari metode dan keakuratan yang beragam.

Peramalan adalah sebuah proses yang digunakan untuk mempelajari suatu permasalahan dalam kondisi tertentu yang nantinya membantu dalam hal pengambilan keputusan dengan alternatif terbaik. Pada dasarnya peramalan mempunyai fungsi sebagai memperkirakan apa yang akan terjadi.

**Kata Kunci :** Peramalan Time Series, Fuzzy clustering, Fuzzy C-Means

## 1. Pendahuluan

Sektor ekonomi merupakan salah satu sektor yang paling dekat dengan manusia, karena kehidupan manusia tak luput dari kegiatan-kegiatan ekonomi seperti transaksi jual-beli. Salah satu bentuk yang pada saat ini trend dan digemari adalah pasar modal atau yang biasa disebut sebagai bursa saham. Saham sendiri merupakan suatu aset yang mudah untuk diperjual belikan dibandingkan dengan aset konvensional seperti tabungan, rumah, obligasi dll. Akan tetapi pergerakan nilai dari suatu saham adalah dinamis dan tidak berbentuk linear sehingga diperlukan metode khusus untuk memodelkannya secara matematis. Peramalan harga saham sangat bermanfaat bagi para pelaku perdagangan saham, dimana keuntungan dari perdagangan saham tersebut sama besar dengan resiko kerugian yang diambil dalam perdagangan saham tersebut. Maka dengan adanya peramalan harga saham yang akurat, diharapkan resiko yang diambil oleh pelaku presentasinya akan berkurang.

Peramalan secara teknik banyak digunakan dalam proses pengambilan keputusan yang tidak lain adalah sebagai representasi dari nilai dari kondisi tertentu yang akan terjadi sehingga menjadi bahan pertimbangan untuk proses pengambilan keputusan sehingga diharapkan alternatif terbaik akan didapatkan.

Peramalan time series merupakan model peramalan dengan input menggunakan data-data historikal yang telah ada, konkret dan mempunyai dependensi atau mempunyai relasi terhadap suatu objek yang dimana dianggap dapat menggambarkan seluruh faktor – faktor dari suatu objek tersebut dan

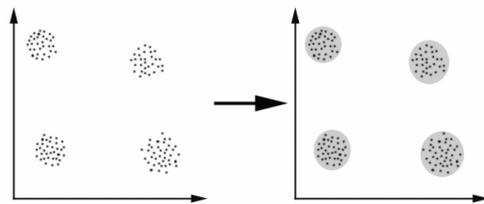
bisa mengabaikan faktor – faktor lain, karena yang dilihat bukanlah hal yang mempengaruhi fluktuasi data, sehingga pertimbangan yang digunakan dalam model time series ini adalah hubungan antar data – data. Peramalan time series juga banyak digunakan antara lain seperti peramalan cuaca, prediksi penjualan suatu produk, data finansial dan lain – lain.

Metode yang digunakan dalam tugas akhir ini adalah Fuzzy clustering karena metode ini dapat mengelompokkan data-data yang “abu-abu” sehingga diharapkan hasil dari proses peramalan akan lebih akurat.

## 2. Data Clustering

Data clustering adalah proses pengelompokan data dalam kelas-kelas atau cluster-cluster sehingga data dalam suatu cluster memiliki tingkat persamaan yang tinggi satu dengan lainnya tetapi sangat berbeda dengan data dalam cluster lain [4]. Jadi, sebuah cluster merupakan kumpulan data-data yang memiliki kesamaan diantara data-data itu sendiri dan ketidaksamaan pada data-data yang terdapat pada kelas lain.

Kita dapat mendeskripsikan contoh sebuah data clustering dengan gambar di bawah ini



Gambar 2.1. Data Clustering dengan Hard Clustering

### 2.1 Fuzzy Clustering

Fuzzy clustering merupakan pengelompokan data yang memiliki karakteristik yang hampir sama secara matematik dalam sebuah kelompok atau kelas. Pada fuzzy clustering, membership function yang akan digunakan, dimodelkan dari data-data yang telah ada. Proses pembentukan membership function ini disebut modelling. Untuk memodelkan data yang non-linear, metode fuzzy clustering yang biasa digunakan adalah C-Means[1].

#### 2.1.1 Algoritma Fuzzy C-Means

Fuzzy C-means Clustering (FCM), atau dikenal juga sebagai Fuzzy ISODATA,

merupakan salah satu metode clustering yang merupakan bagian dari metode Hard K-Means. FCM menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kelas atau cluster terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1 [4].

Konsep dasar FCM pertama kali adalah menentukan pusat cluster pada kondisi awal pusat cluster ini masih belum akurat. Setiap data memiliki derajat keanggotaan untuk tiap cluster dengan cara memperbaiki pusat cluster dan nilai keanggotaan tiap data secara berulang, maka akan dapat dilihat bahwa pusat cluster akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimisasi fungsi obyektif.

Algoritma dari Fuzzy C-Means[2];

1. Input data yang akan di cluster X, berupa matriks berukuran n x m (n=jumlah sampel data, m=attribute setiap data).
2. Tentukan :
  - o Jumlah cluster = c;
  - o Pangkat = w;
  - o Maksimum iterasi = MaxIter;
  - o Error terkecil yang diharapkan =  $\epsilon$ ;
  - o Fungsi obyektif awal =  $P_0 = 0$ ;
  - o Iterasi awal =  $t = 1$ ;
3. Bangkitkan bilangan random  $U_{ik}$ , dimana  $i=1,2,3,\dots,n$ ;  $k=1,2,\dots,c$ ; sebagai elemen-elemen matriks partisi awal U.

$$U_{ik} = \begin{pmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1k} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2k} \\ \dots & \dots & \dots & \dots \\ \mu_{i1} & \mu_{i2} & \dots & \mu_{ik} \end{pmatrix} \quad (1)$$

4. Hitung pusat cluster ke-k:  $V_{kj}$ , dengan  $k=1,2,\dots,c$ ; dan  $j=1,2,\dots,m$ .

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w * X_{ij}}{\sum_{i=1}^n \mu_{ik}} \quad (2)$$

5. Hitung fungsi obyektif pada iterasi ke-t,  $P_t$  :

$$P_t = \sum_{j=1}^m \sum_{k=1}^c \left[ \left( \sum_{i=1}^n (X_{ij} - V_{kj})^2 \right) (\mu_{ik})^w \right] \quad (3)$$

6. Hitung perubahan matriks partisi, dengan :  $i=1,2,\dots,n$ ; dan  $k=1,2,\dots,c$ :

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}$$

- (4)
7. Cek kondisi berhenti :
- Jika :  $(|P_t - P_{t-1}| < \epsilon)$  atau  $(t > \text{maxIter})$  maka berhenti;
  - Jika tidak :  $t=t+1$ , ulangi langkah ke-4.

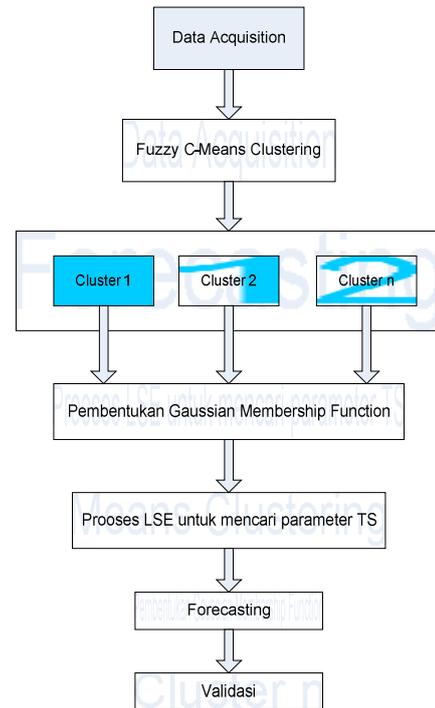
## 2.2 Least Square Error (LSE)

Setelah fuzzy membership function terbentuk, langkah berikutnya adalah proses defuzzifikasi, karena fuzzy inference system yang digunakan adalah tipe Takagi-Sugeno, proses defuzzifikasi dapat dilakukan setelah parameter Takagi-Sugeno  $\theta$  diketahui nilainya. Perhitungan nilai  $\theta$  untuk MISO (Multi Input Single Output) inference system dilakukan dengan langkah-langkah berikut [3];

- Menghitung  $\beta$ . Nilai  $\mu_G$  didapat dengan memasukkan input ke Gaussian Membership Function (GMF) yang bersesuaian.
- Menghitung  $\gamma$
- Membentuk matrik  $XIe$  yang berisi kumpulan konstanta 1 dan n-input
$$XIe = [1, x_{1\_s}, x_{2\_s}, \dots, x_{n\_s}]$$
- Parameter  $\theta$  dapat dinotasikan dengan
$$[\theta] = ([Y \cdot XIe]^T \cdot [Y \cdot XIe])^{-1} \cdot [Y \cdot XIe]^T \cdot [Y]$$

## 3. Arsitektur Sistem

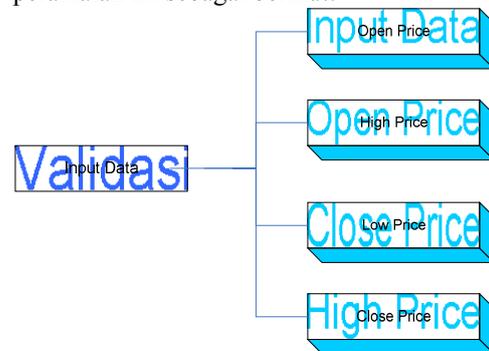
Tujuan utama dari pembuatan fuzzy clustering ini adalah memodelkan data time-series saham yang non-linear ke model fuzzy. Model fuzzy ini akan digunakan untuk melakukan forecasting data.



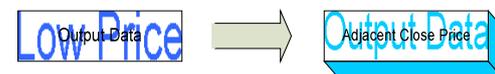
Gambar 3.1 Arsitektur Sistem

## 3.1 Input dan Output data

Input data dan output data pada peramalan ini sebagai berikut:



Gambar 3.2 Input data saham



Gambar 3.3 Output data saham

## 3.2 Implementasi Algoritma FCM

Yaitu berisi pembuatan program Fuzzy C-means sesuai algoritma pada subbab 2.1.1.

### 3.3 Pembentukan Gaussian Membership Function

Untuk setiap cluster dari hasil proses Fuzzy C-Means akan dibuat Gaussian Membership Function-nya. Bentuk Gaussian Membership Function dari matrik U. Dari proses Fuzzy C-Means telah didapatkan center dari cluster V yang juga merupakan center daripada Gaussian Membership Function. Rumus daripada fungsi GMF adalah sebagai berikut :

$$f(x, \sigma, c) = e^{\frac{-0.5(x-c)^2}{\sigma^2}} \quad (5)$$

### 3.4 Proses Least Square Error

Proses ini menggunakan sistem inference Takagi-Sugeno. Tahap-tahap LSE:

1. Bentuk matrik Xie

Matrik Xie adalah matrik yang berisi nilai 1 pada kolom pertama dan berisi nilai input pada kolom 2 hingga kolom atribut terakhir

2. Hitung degree of fulfilment ( $\beta$ )

Degree of fulfilment dihitung menggunakan product operator yaitu mengalikan setiap nilai degree of membership dari nilai input terhadap Gaussian Membership Function-nya

3. Normalisasi degree of fulfilment ( $\gamma$ )

Setelah mendapat nilai dari degree of fulfilment, hitung nilai normalisasi dari degree of fulfilment

4. Menghitung nilai Consequent parameter Takagi-Sugeno ( $\theta$ )

Setelah nilai  $\gamma$  didapat maka nilai dariconsequent parameter .

### 3.5 Forecasting

Setelah consequent parameter Takagi-Sugeno didapatkan, maka dapat dilakukan forecasting menggunakan sistem fuzzy yang telah terbentuk. Proses ini sebenarnya hampir sama dengan proses LSE, hanya saja yang membedakan adalah data yang digunakan. Jika pada proses LSE data yang digunakan adalah data training, sedangkan pada proses forecasting data yang digunakan adalah data test.

Untuk melakukan forecasting dapat dilakukan dengan langkah sebagai berikut:

1. Masukkan nilai input ke dalam matrik Xie
2. Hitung nilai  $\beta$
3. Hitung nilai  $\gamma$
4. Hitung matrik [ $\gamma$  Xie]
5. Hitung output hasil forecast [ $\hat{y}$ ] = [ $\gamma$  Xie][ $\theta$ ] dengan menggunakan  $\theta$  yang telah didapat dari perhitungan LSE

### 3.6 Validasi

Setelah proses forecasting dilakukan maka proses berikutnya adalah melihat error yang terjadi pada hasil forecast. Nilai error dapat dihitung dengan rumus sebagai berikut :

$$\varepsilon = \hat{x} - x \quad (6)$$

## 4. Uji Coba dan Analisa

Pada bab ini akan dibahas mengenai uji coba dan analisa data time series saham yang mempunyai karakteristik random dan non stationer. Data saham merupakan data yang sulit untuk diprediksi karena pergerakan harga saham yang memang sulit untuk menentukan polanya. Kesepuluh data saham yang dipilih dalam peramalan ini, yaitu:

1. Bumi Resources Tbk (BUMI.JK)
2. Astra International Tbk (ASII.JK)
3. Telekomunikasi Indonesia (TLKM.JK)
4. Indosat Tbk PT (IIT)
5. Bank International Indonesia Tbk (BNII.JK).
6. Bank Central Asia Tbk (BBCA)
7. Bank Rakyat Indonesia Tbk (BBRI)
8. Bank Mandiri Tbk (BMRI)
9. Unilever Indonesia Tbk (UNVR)
10. Indofood Sukses Makmur Tbk (INDF).

Dalam proses analisa, data dari kesepuluh perusahaan ini mempunyai rentan waktu sampai tanggal 1 Januari 2010. Untuk data BUMI mempunyai 1772, ASII mempunyai 2326 data, TLKM mempunyai 1288 data, IIT mempunyai 2510 data, BNII mempunyai 1251 data, BBCA mempunyai 1410 data, BBRI mempunyai 1527 data, BMRI mempunyai 1612 data, UNVR

mempunyai 1575 data dan INDF mempunyai 2335 data.

Proses analisa ini bertujuan untuk mengetahui optimasi dari metode Fuzzy C-Means. Dalam proses ini akan diuji parameter dari Fuzzy C-Means yaitu jumlah cluster, tingkat fuzziness dan jumlah data testing yang digunakan dalam melakukan proses peramalan.

**Tabel 4.1** Hasil Peramalan

Nama Perusahaan	MSE	Tipe Data
BUMI.JK	1.83E-05	Trend
ASII.JK	7.96E-04	Trend
TLKM.JK	0.0014746	Trend
IIT	1.91E-04	Random
BNIL.JK	7.42E-09	Random
BBCA.JK	0.1415623	Trend
BBRI.JK	3.21E-04	Trend
BMRI.JK	0.0013104	Random
UNVR.JK	3.71E-08	Random
INDF.JK	6.89E-05	Random

Dari tabel di atas, 4 dari 5 nilai MSE terkecil didapatkan dari tipe data berjenis "Random". Sehingga metode fuzzy C-Means ini sangat sesuai digunakan untuk tipe data random.

## 5. Kesimpulan

Dari hasil uji coba mengenai peramalan time series harga saham menggunakan feed forward neural network dapat diambil beberapa kesimpulan, yaitu:

- ❖ Tiap-tiap data perusahaan mempunyai tipe yang berbeda dan membutuhkan nilai parameter yang berbeda untuk menghasilkan error yang terkecil.
- ❖ Semakin banyak jumlah cluster tidak menjamin semakin baiknya hasil peramalan.
- ❖ Tipe data training sangat berpengaruh terhadap akurasi hasil peramalan.
- ❖ Kemampuan metode Fuzzy C-Means cukup baik dalam melakukan peramalan untuk tipe data random.

## DAFTAR PUSTAKA

- [1]. Chandra Lindawati, "Fuzzy Clustering dengan Metode Gustafson-Kessel untuk Forecasting Data Electrical Load", Universitas Kristen Petra, 2007.
- [2]. Irawan M. Isa, Satriyanto Edi, "Virtual Pointer Untuk Identifikasi Isyarat Tangan Sebagai Pengendali Gerakan Robot Secara Real-Time", Bidang Ilmu Komputer- jur. Matematika FMIPA - ITS, Politeknik Elektronika Negeri Surabaya (PENS), Surabaya, 2008.
- [3]. Pasila Felix, Gunawan Dharma, Ferdinando Hany, "Evolutionary Algorithm pada Fuzzy Clustering Systems Metode Gustafson-Kessel untuk Forecasting Electrical Load Data Time-Series", Electrical Engineering Department, Petra Christian University Surabaya, 2008.
- [4]. Taufiq Luthfi Emha, "Fuzzy C-Means Untuk Clustering Data (Studi Kasus : Data Performance Mengajar Dosen)", STMIK AMIKOM Yogyakarta, 2007.
- [5]. Wijaya Lie Handra, "Fuzzy Clustering C-Means Method for Electrical Load Time-Series Data Forecasting ", Petra Christian University, 2007.