

Studi Kasus

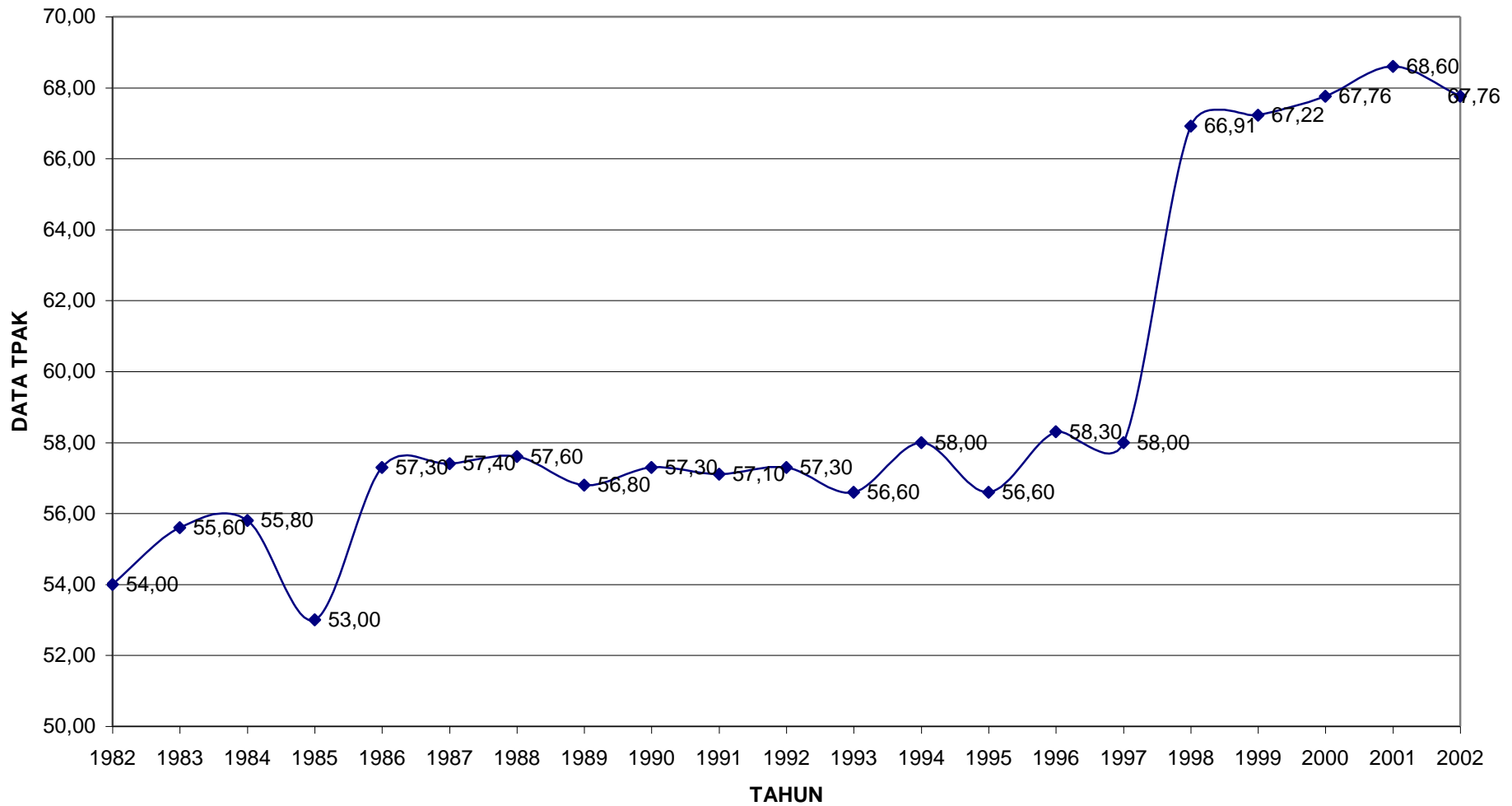
Dr. Suyanto, S.T., M.Sc.
HP/WA: 0812 845 12345

Intelligence Computing Multimedia (ICM)
Informatics faculty – Telkom University

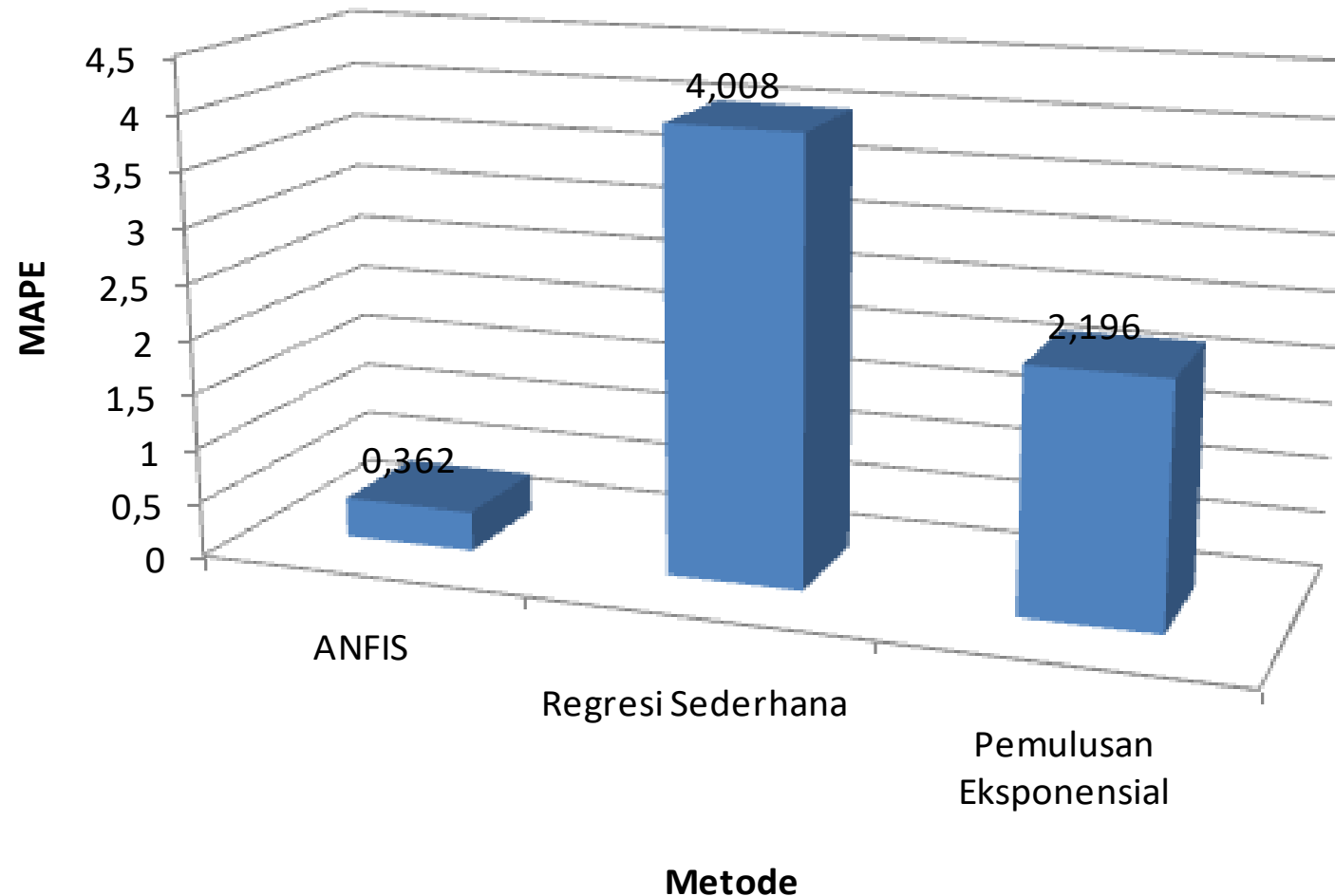
Peramalan Tingkat Partisipasi Angkatan Kerja

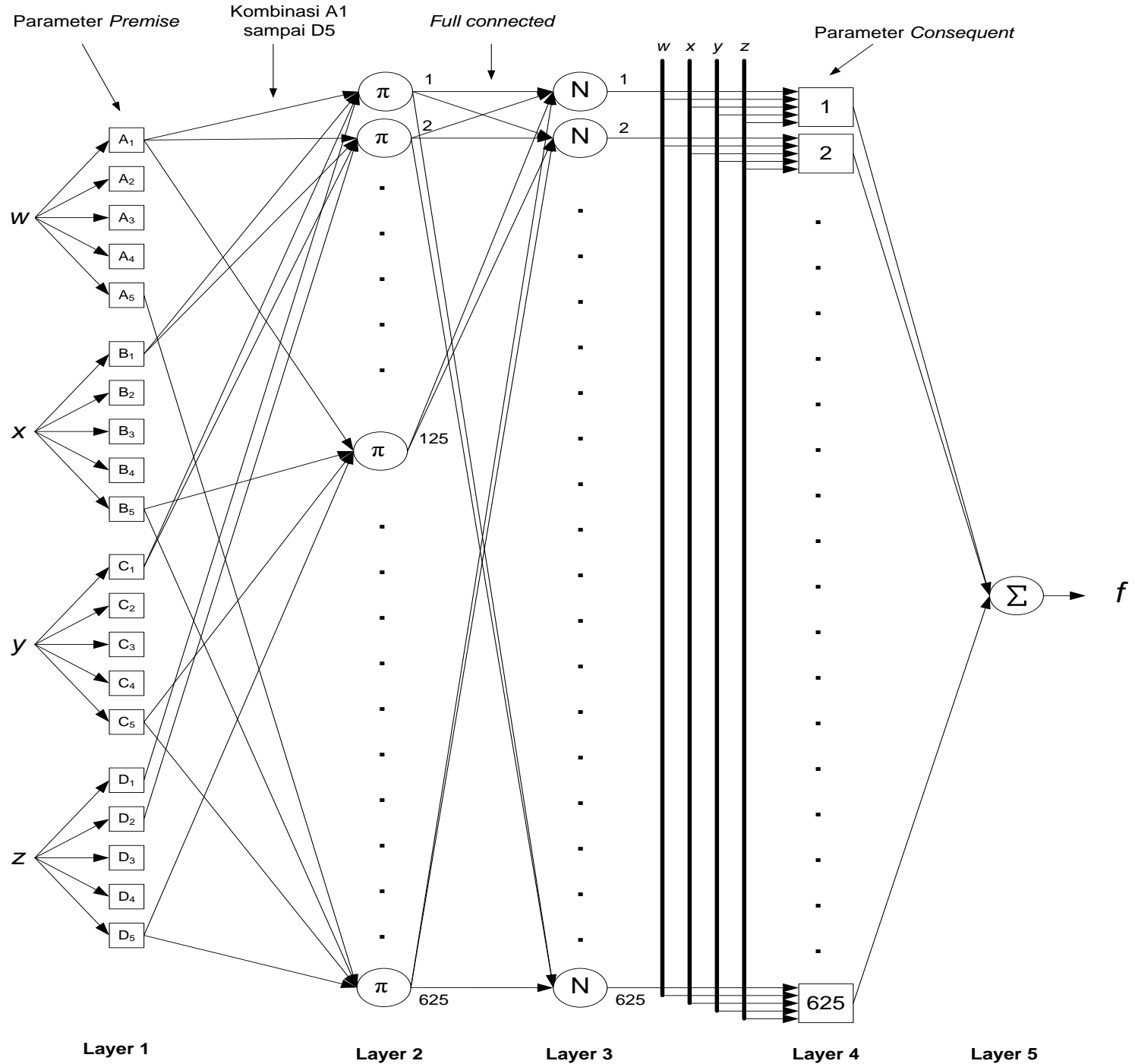
- Peramalan Data Time Series
- ANFIS

GRAFIK POLA DATA TPAK NASIONAL PERIODE 1982-2002



MAPE untuk ANFIS, Regresi Sederhana, dan Pemulusan Eksponensial

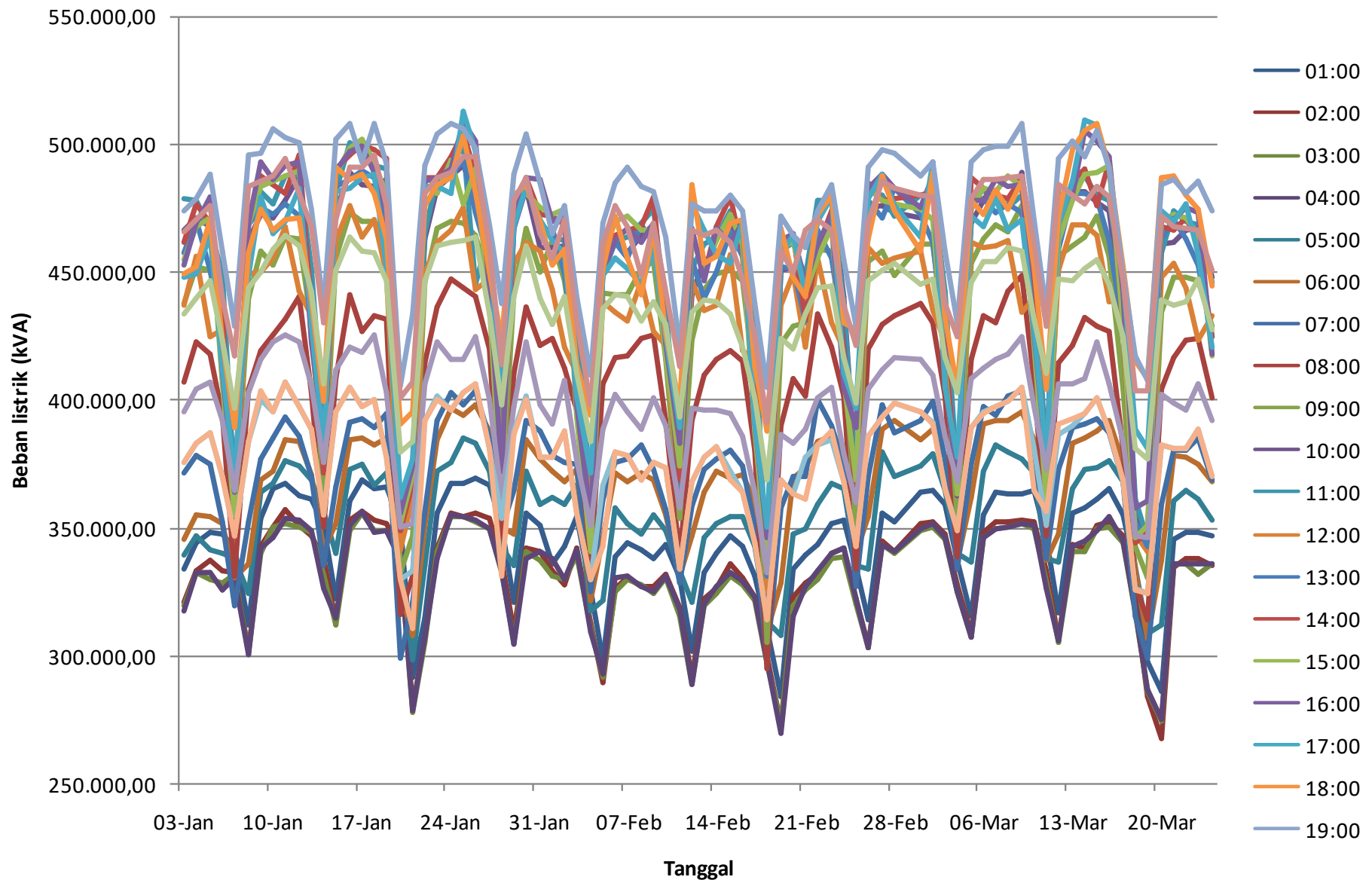




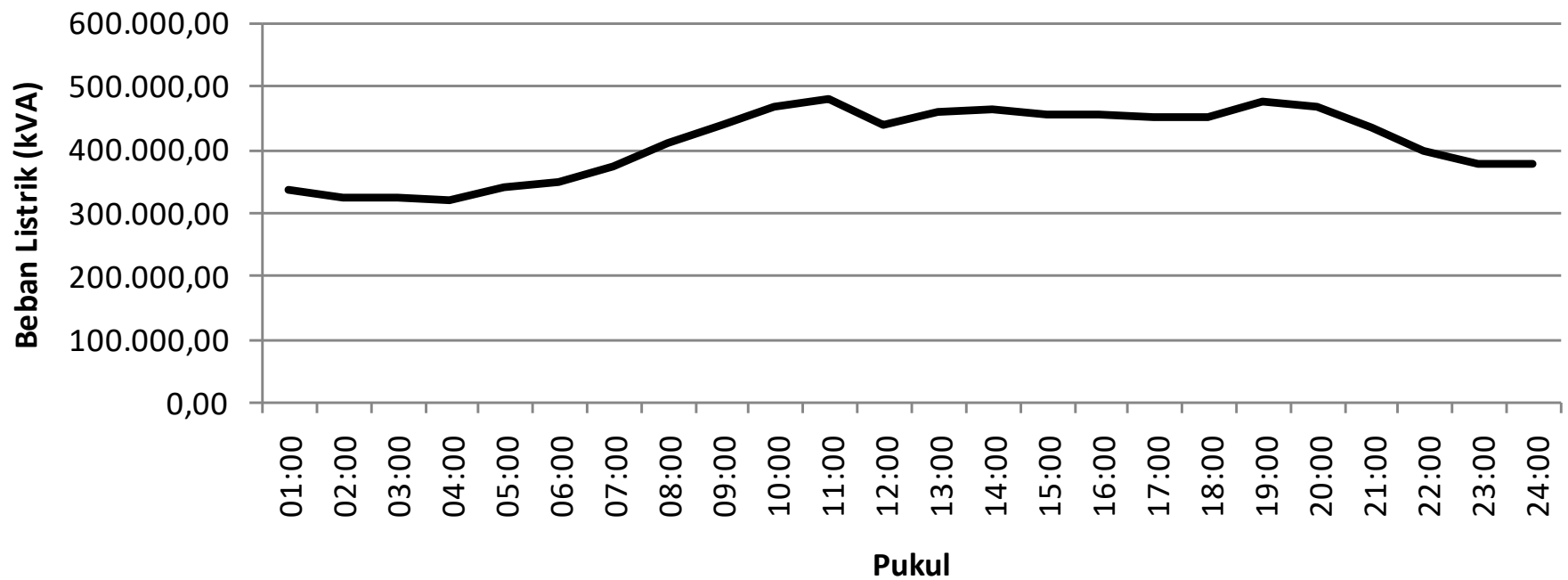
Prediksi Beban Listrik Jangka Pendek

- Peramalan Data Time Series
- NEFPROX

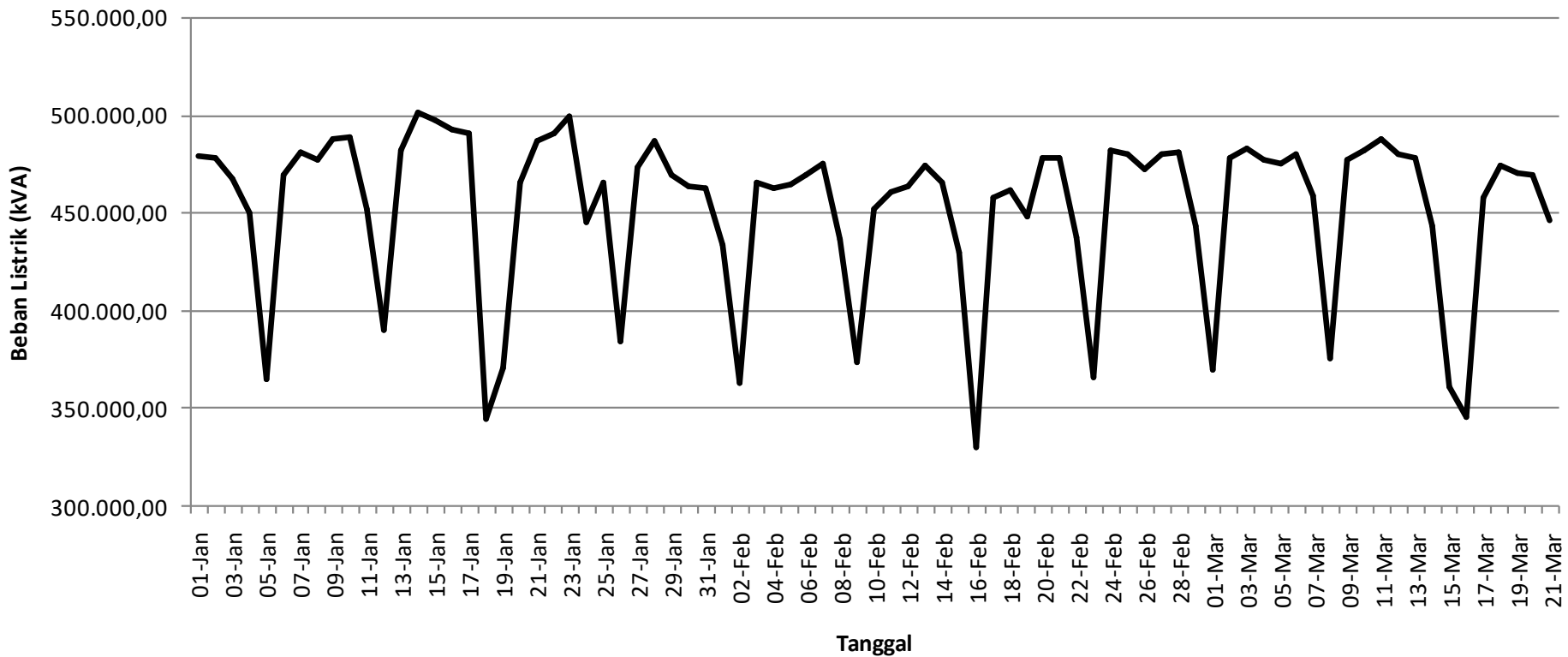
Data Beban Listrik per jam (pukul 01:00 - 24:00) Periode 03 Januari - 24 Maret 2007 di suatu APJ



Beban listrik pada tanggal 03 Jan 2007



Beban Listrik pada pukul 11:00 untuk Periode 03 Jan - 24 Mar 2007



Parameter NEFPROX

- *fuzzy set* yang digunakan adalah segitiga
- *Jumlah fuzzy set* = 4 atau 5
- *Jumlah input* = 4 atau 5
- *Learning rate* (σ) = 0,0025; 0,006; dan 0,01
- Fungsi kenggotaan memiliki *intersect* 0,5

Domain fungsi keanggotaan untuk data per jam

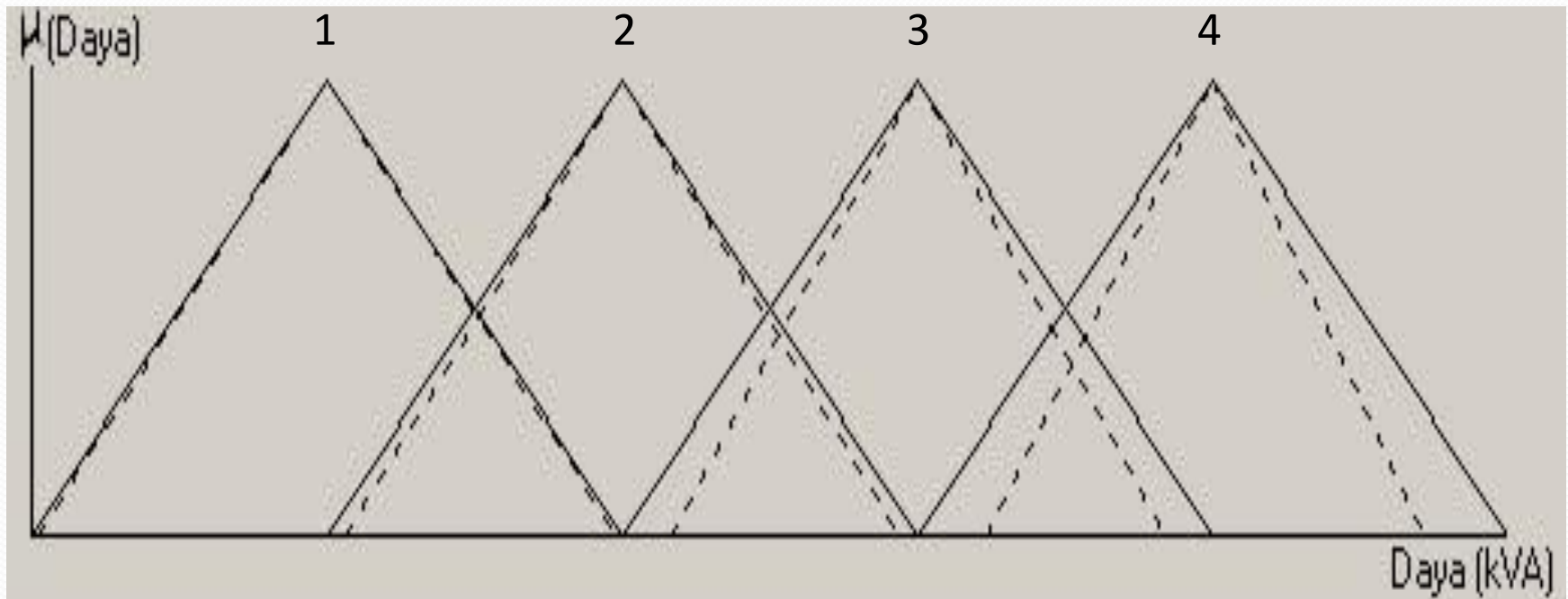
Pukul	Domain (kVA)	Pukul	Domain (kVA)
01:00	270000 - 390000	13:00	320000 - 510000
02:00	250000 - 370000	14:00	320000 - 535000
03:00	250000 - 370000	15:00	320000 - 535000
04:00	250000 - 370000	16:00	330000 - 530000
05:00	280000 - 415000	17:00	340000 - 540000
06:00	275000 - 420000	18:00	375000 - 525000
07:00	275000 - 420000	19:00	375000 - 530000
08:00	280000 - 475000	20:00	370000 - 510000
09:00	290000 - 500000	21:00	345000 - 475000
10:00	310000 - 510000	22:00	320000 - 440000
11:00	310000 - 520000	23:00	295000 - 435000
12:00	300000 - 500000	24:00	300000 - 420000

Structure *Learning*

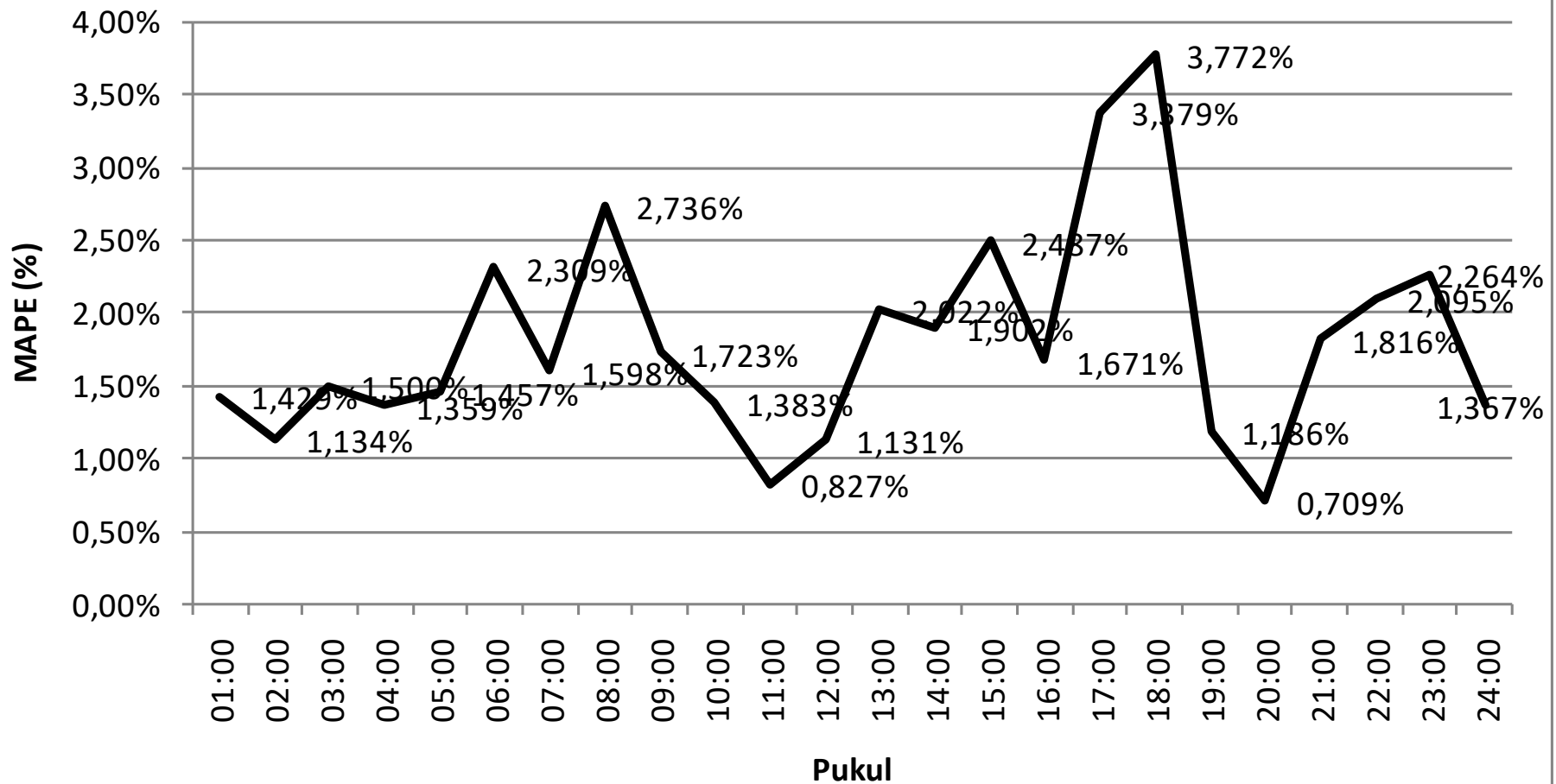
- Untuk menemukan struktur (arsitektur) NEFPROX yang paling optimal.
- Jika *jumlah fuzzy set* (F) = 4 dan *jumlah input* (P) = 4, maka total aturan (*fuzzy rule*) yang harus dibuat pada NEFPROX adalah $F^P = 4^4 = 256$ aturan.
- Jika *jumlah fuzzy set* (F) = 5 dan *jumlah input* (P) = 5, maka total aturan harus dibuat pada NEFPROX adalah $F^P = 5^5 = 3125$ aturan.

Pukul	Jumlah <i>input</i> optimal	Jumlah <i>Fuzzy Set</i> optimal	σ	Jumlah Aturan yang dihasilkan
01:00	5	4	0.01	40
02:00	5	5	0.006	36
03:00	5	5	0.01	39
04:00	5	5	0.01	43
05:00	5	5	0.01	42
06:00	5	5	0.01	44
07:00	5	5	0.006	49
08:00	5	5	0.01	43
09:00	5	5	0.006	35
10:00	5	5	0.01	35
11:00	5	5	0.006	36
12:00	5	5	0.006	38
13:00	4	5	0.0025	24
14:00	5	5	0.006	45
15:00	5	4	0.0025	43
16:00	5	4	0.0025	35
17:00	4	4	0.006	30
18:00	5	4	0.0025	51
19:00	5	4	0.006	34
20:00	5	4	0.006	43
21:00	5	4	0.006	35
22:00	4	4	0.006	36
23:00	4	4	0.006	34
24:00	4	5	0.01	45

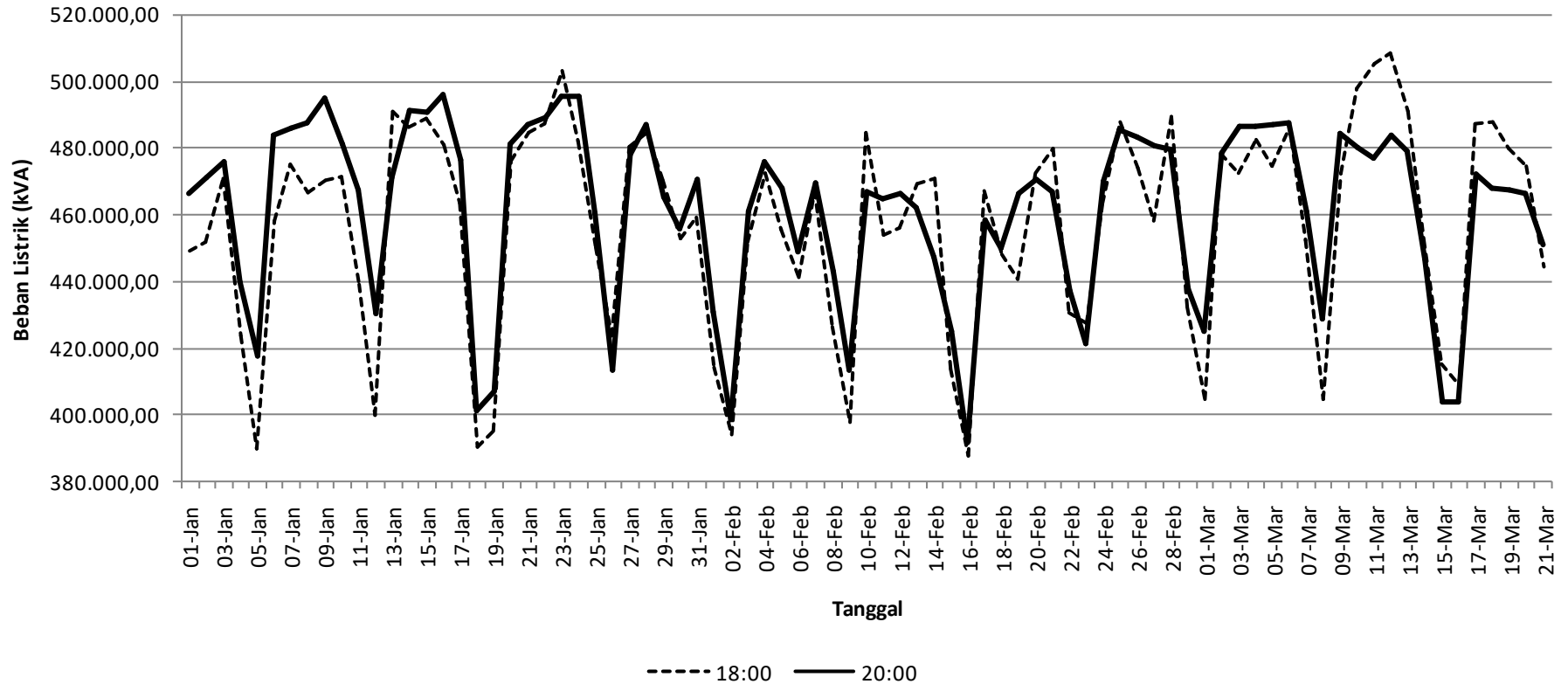
Parameter *Learning*



MAPE untuk data *Test Set*: 11-17 Mar 2007



Beban Listrik pada pukul 18:00 dan 20:00



Klasifikasi Data dengan *Imbalance Class*

- Bagaimana mengenali pelanggan yang potensial terhadap *couponing* oleh toko.
- Kupon adalah voucher untuk promosi yang diberikan pada pelanggan melalui berbagai macam media.
- Seorang pelanggan yang membeli barang yang diiklankan di kupon, akan mendapatkan diskon sesuai yang tertera pada kupon.
- Selama ini, kupon disebarakan melalui majalah atau diberikan secara langsung kepada pelanggan ketika berada di kasir akan melakukan pembayaran.

Data

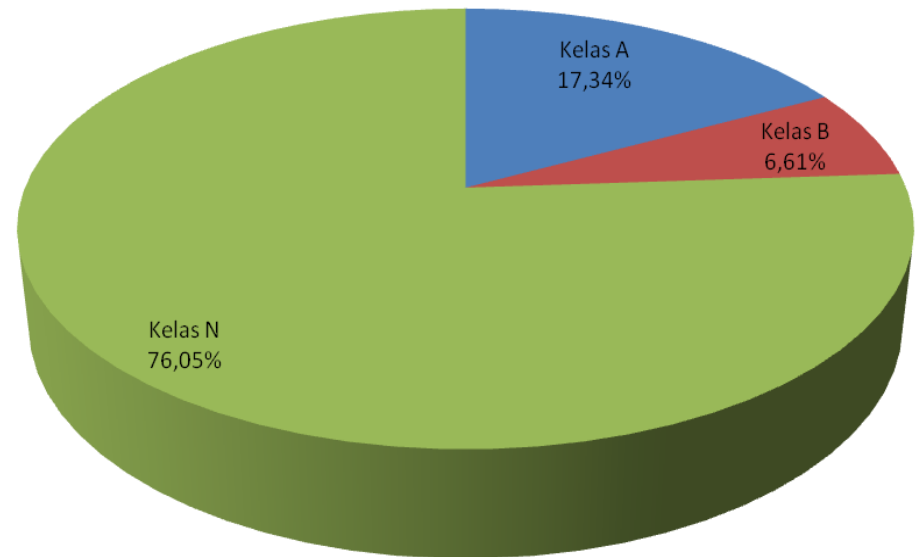
- Data-data untuk pelatihan (*training set*) dan pengujian (*test set*) berupa perilaku penebusan kupon oleh pelanggan.
- Pada data tersebut, pelanggan dibagi menjadi tiga kelas yaitu A, B, dan N.
- Setiap record memiliki 22 atribut.

Data

No	Nama Atribut	Keterangan
1	ID	Nomor pelanggan
2	C10001	Jumlah penebusan kupon C10001 oleh pelanggan
3	C10002	Jumlah penebusan kupon C10002 oleh pelanggan
...
21	C10020	Jumlah penebusan kupon C10020 oleh pelanggan
22	COUPON	atribut target

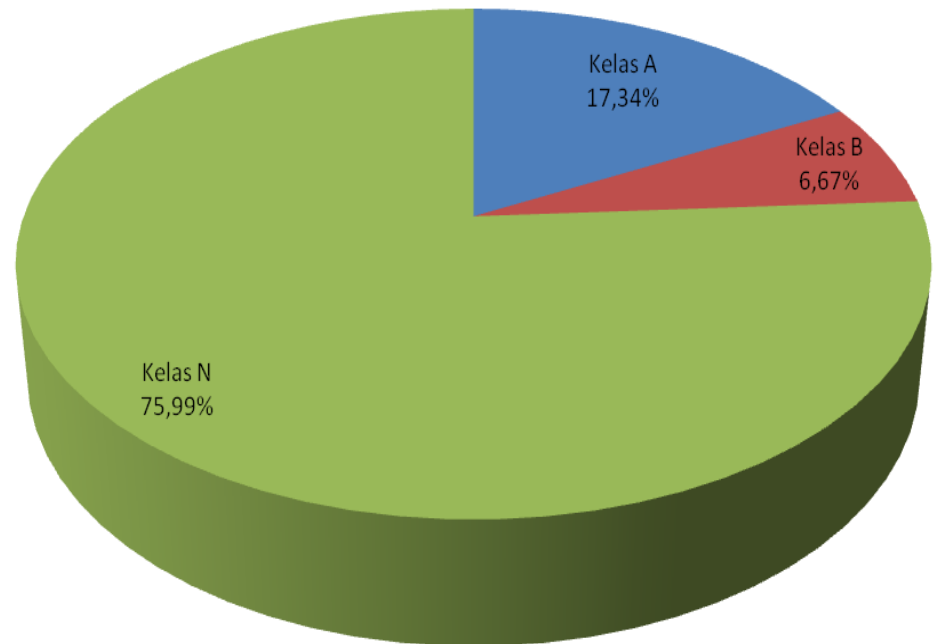
Training Set

- 50 ribu records
- Kelas A berjumlah 8668 *record* (17,34%)
- Kelas B berjumlah 3307 *record* (6,61%)
- Kelas N berjumlah 38025 *record* (76,05%)



Test Set

- 50 ribu records
- Kelas A berjumlah 8668 *record* (17,34%)
- Kelas B berjumlah 3335 *record* (6,67%)
- Kelas N berjumlah 37997 *record* (75,99%)



Prediksi Churn Pelanggan Telepon

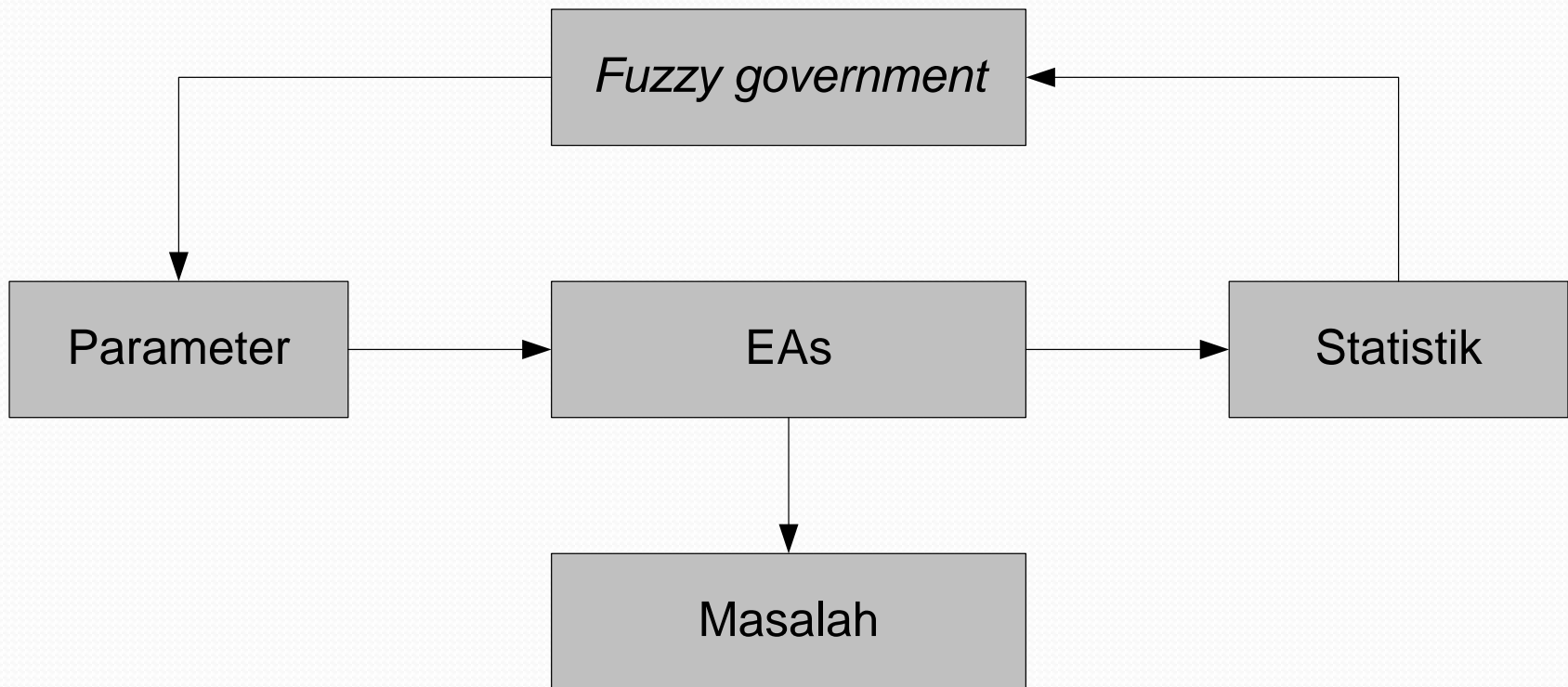
- Input: Data Pelanggan
 - Profil
 - Perilaku pemakaian telepon, dsb.
- Output:
 - Daftar pelanggan yang potensial *churn*

Churn

- *Churn*: pindahnya pelanggan dari satu operator ke operator lain.
- *Voluntary churn*: pindah dengan kemauan sendiri
 - karena kompetitor menawarkan produk yang lebih murah dan berkualitas
- *Involuntary* adalah *churn* bukan karena kemauan sendiri
 - karena pelanggan sudah lama tidak membayar tagihannya sehingga menyebabkan dicabutnya layanan terhadap pelanggan tersebut.

Atribut	Keterangan
A1	Kategori Cluster yang didefinisikan oleh carrier
A2	Jenis layanan yang dipakai pelanggan
A3	Jenis Kelamin
A4	Kategori tanggal pembayaran tagihan untuk bulan tagihan M2
A5	Kategori frekuensi pemanggilan dari Domestic to Lokal
A6	Kategori frekuensi pemanggilan dari Domestic to PSTN
A7	Kategori frekuensi pemanggilan dari Lokal to Flexy
A8	Kategori frekuensi pemanggilan dari Lokal to OLO
A9	Kategori frekuensi pemanggilan dari Domestic to PSTN
A10	Kategori frekuensi pemanggilan SMO
A11	Kategori frekuensi pemanggilan SMP
A12	Kategori durasi lamanya nomor telepon tersebut dipakai sampai bulan ke M2
A13	Kategori durasi lamanya nomor telepon tersebut dipakai dr awal
A14	Kategori jumlah revenue dari pemakaian layanan oleh pelanggan pada bulan M2
A15	Trend revenue bulan M2 terhadap bulan M3
A16	Pengkategorian SDEV dari payment
A17	Kombinasi dari variabel kategori panggilan domestic ke seluler (D2C) dan panggilan domestic ke PSTN (D2P) untuk bulan M2
A18	Kombinasi dari variabel kategori panggilan lokal ke flexi (L2F), lokal ke OLO (L2O) dan panggilan lokal ke PSTN (L2P) untuk bulan M2
A19	Kombinasi kategori penggunaan voice per hari selama satu minggu (Minggu - Sabtu) di bulan M2
A20	Kombinasi variabel TREND1-TREND4
A21	Persentase durasi pemakaian jenis voice ke jaringan Telkom dibandingkan seluruh durasi pemakaian jenis voice
A22	Label Kelas (Churn / Active)

Adaptive EAs



Contoh Kasus

Kota	Tipe	Tipe Pembayaran	Tagihan Bulanan	Jumlah Panggilan	Panggilan Tidak Normal	Churn
Jakarta	Pemerintah	Cash	Besar	10	Sedikit	Tidak
Jakarta	Corporate	Kartu Kredit	Sedang	8	Sedang	Tidak
Jakarta	Corporate	Kartu Kredit	Kecil	5	Banyak	Tidak
Surabaya	Corporate	Cash	Kecil	3	Banyak	Ya
Surabaya	Corporate	Cash	Kecil	2	Banyak	Ya
Surabaya	Corporate	Kartu Kredit	Besar	1	Sedang	Tidak
Jakarta	Corporate	Kartu Kredit	Sedang	9	Sedang	Tidak
Jakarta	Corporate	Kartu Kredit	Sedang	7	Sedang	Tidak
Jakarta	Corporate	Kartu Kredit	Sedang	6	Sedang	Tidak
Jakarta	Corporate	Kartu Kredit	Sedang	4	Sedang	Tidak
Jakarta	Corporate	Kartu Kredit	Sedang	9	Sedang	Tidak

Diskritisasi dan transformasi

Kota	Tipe	Tipe Pembayaran	Tagihan Bulanan	Jumlah Panggilan	Panggilan TidakNormal	Churn
Surabaya	Corporate	Cash	Kecil	[2...3]	Banyak	Class1
Surabaya	Corporate	Cash	Kecil	[1...2]	Banyak	Class1
Jakarta	Corporate	Kartu Kredit	Sedang	[3...4]	Sedang	Class2
Jakarta	Corporate	Kartu Kredit	Sedang	[5...6]	Sedang	Class2
Jakarta	Corporate	Kartu Kredit	Kecil	[4...5]	Banyak	Class2
Jakarta	Corporate	Kartu Kredit	Sedang	[7...8]	Sedang	Class2
Surabaya	Corporate	Kartu Kredit	Besar	[0...1]	Sedang	Class2
Jakarta	Corporate	Kartu Kredit	Sedang	[8...9]	Sedang	Class2
Jakarta	Corporate	Kartu Kredit	Sedang	[6...7]	Sedang	Class2
Jakarta	Pemerintah	Cash	Besar	[9...10]	Sedikit	Class2
Surabaya	Corporate	Cash	Kecil	[2...3]	Banyak	Class1

APACS

- APACS: *Automatic Analysis And Classification of Conceptual Patterns*
- APACS adalah merupakan teknik induksi probabilistik.
- Diantara beberapa pasangan atribut yang mungkin, APACS dapat mengidentifikasi pasangan yang memiliki hubungan asosiatif bahkan pada basis data yang memiliki data *noise* dan memiliki banyak nilai yang hilang.

APACS → *First Order Rule*

First Order Rule	Nilai Asosiasi terhadap Churn (d)	Nilai Ketidakpastian Asosiasi/ Bobot (\overline{w})
Kota = 'Surabaya'	2.41	0.30
Kota = 'Jakarta'	-2.41	-1
Tipe Pembayaran = 'Cash'	2.41	0.30
Tipe Pembayaran = 'Kartu Kredit'	-2.41	-1
Tagihan Bulanan = 'Kecil'	2.41	0.30
Jumlah Panggilan = '[2...3]'	2.10	1
Jumlah Panggilan = '[1...2]'	2.10	1
Panggilan Tidaknormal = 'Banyak'	2.41	0.30

Populasi pada *Second Order rule*

TipePembayaran = 'Kartu Kredit'

TagihanBulanan = 'Kecil'

TipePembayaran = 'Kartu Kredit'

Kota = 'Surabaya'

Kota = 'Surabaya'

TagihanBulanan = 'Kecil'

TipePembayaran = 'Kartu Kredit'

TipePembayaran = 'Kartu Kredit'

JumlahPanggilan = 'Sedikit'

Kota = 'Surabaya'

Crossover

O₁ TipePembayaran = 'Kartu Kredit' TagihanBulanan = 'Kecil'

O₂ JumlahPanggilan = 'Sedikit' Kota = 'Surabaya'

A₁ TipePembayaran = 'Kartu Kredit' Kota = 'Surabaya'

A₂ JumlahPanggilan = 'Sedikit' TagihanBulanan = 'Kecil'

Contoh rule hasil learning

Order	Rule	Weight of Evidence
1st Order	kota = 'Surabaya'	0.3
1st Order	tipepembayaran = 'Cash'	0.3
2nd Order	jumlahpanggilan = '[2...3]' AND panggilantidaknormal = 'Banyak'	1.0
2nd Order	panggilantidaknormal = 'Banyak' AND tipepembayaran = 'Cash'	1.0
3th Order	jumlahpanggilan = '[2...3]' AND panggilantidaknormal = 'Banyak' AND tipepembayaran = 'Cash'	1.0
3th Order	jumlahpanggilan = '[2...3]' AND tagihanbulanan = 'Kecil' AND tipepembayaran = 'Cash'	1.0
4th Order	jumlahpanggilan = '[2...3]' AND panggilantidaknormal = 'Banyak' AND tagihanbulanan = 'Kecil' AND tipepembayaran = 'Cash'	1.0

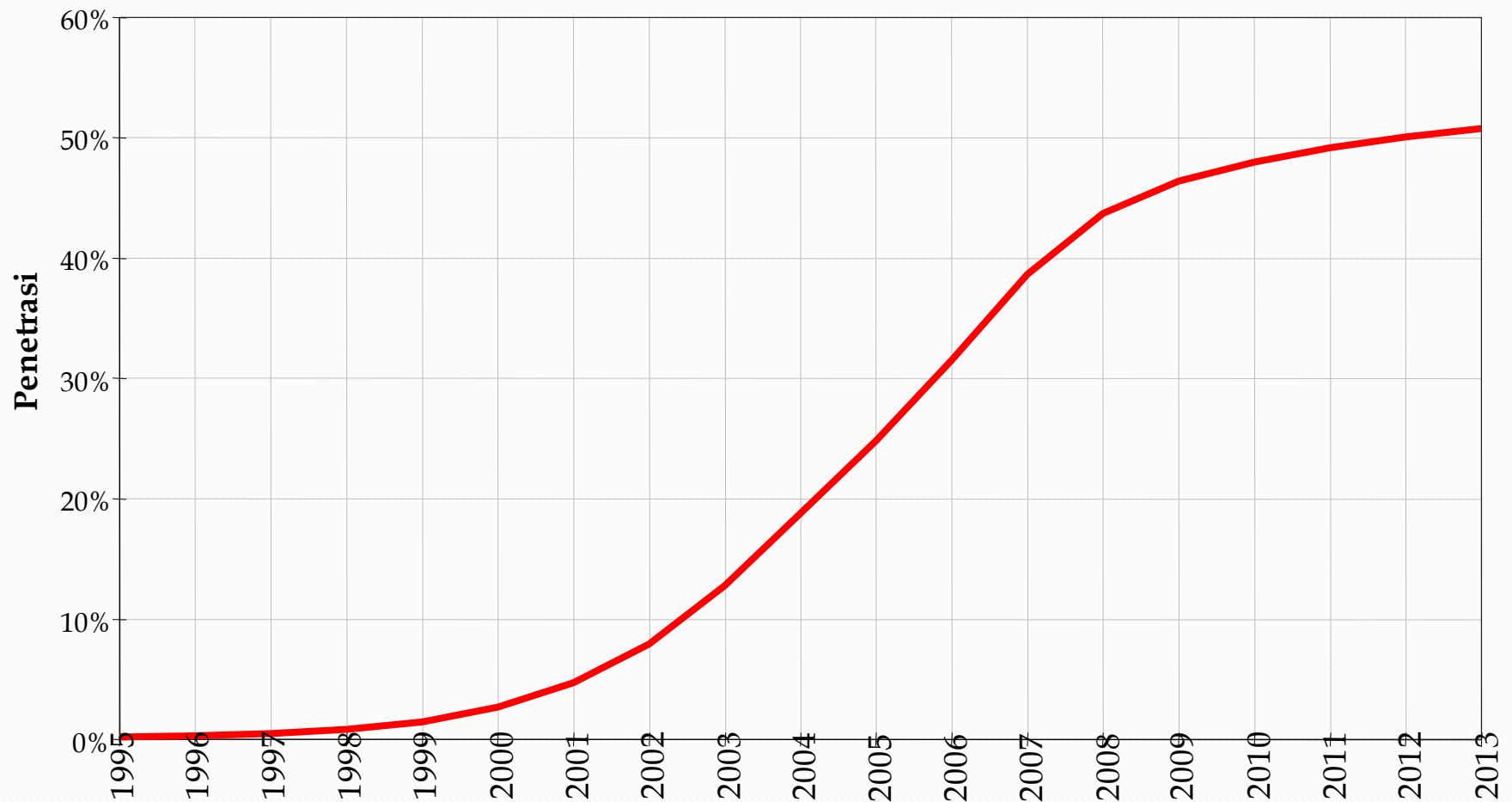
Karakteristik Trafik

- Sangat penting untuk dipahami
- Untuk menentukan dimensi jaringan
- Sistem prediksi berbasis data *time series*

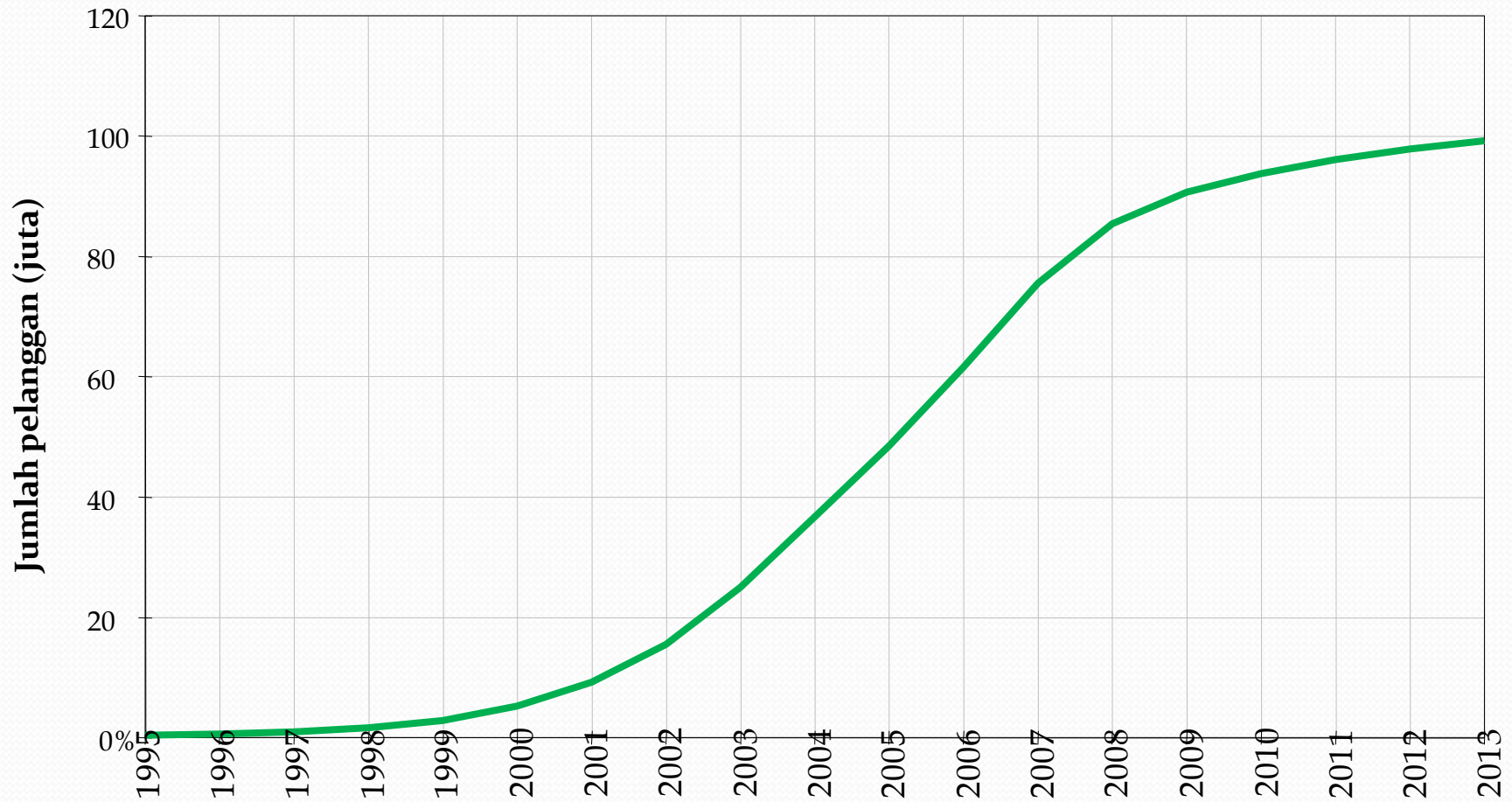
Traffic Forecasting

- Pertumbuhan penetrasi
- Pertumbuhan jumlah pelanggan
- Pertumbuhan trafik

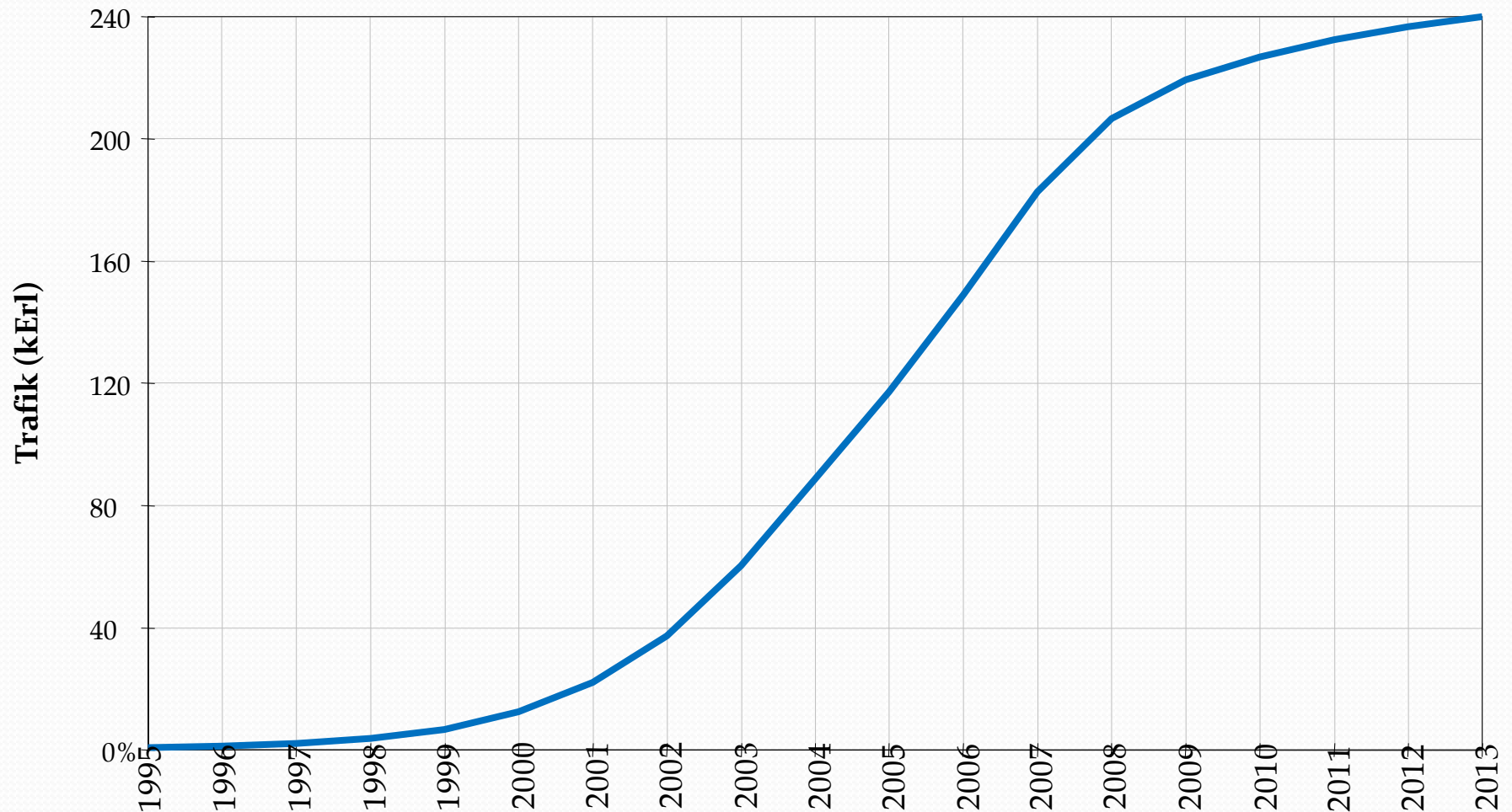
Pertumbuhan Penetrasi



Pertumbuhan jumlah pelanggan

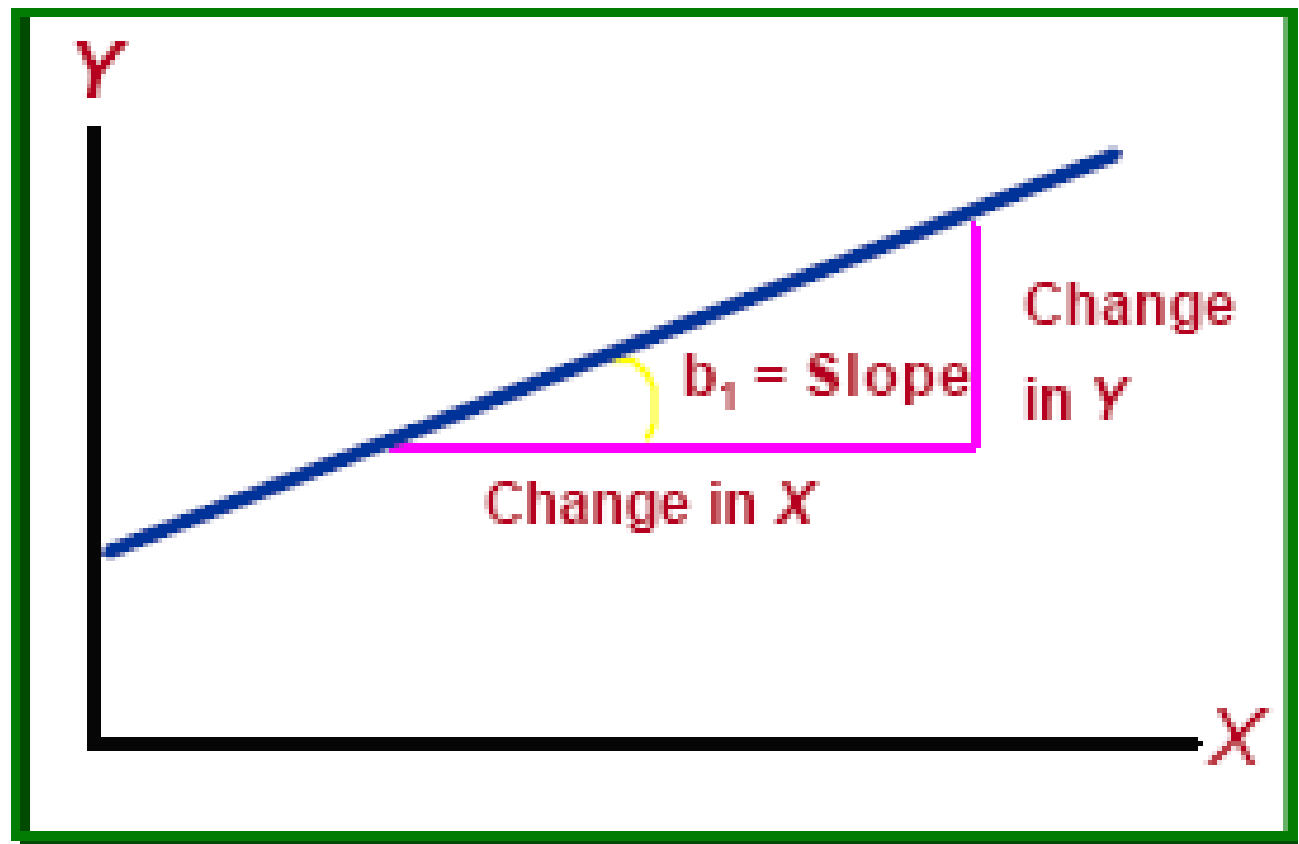


Pertumbuhan trafik



Model Regresi Linear

Linear Model
$$Y_i = a + b_1 X_i + \epsilon_i$$



Prediksi *Time Series*

$$z = a_0 + a_1 y_1 + a_2 y_2 + \dots + a_k y_k$$

- y_1 sampai y_k adalah masukan yang berupa data-data sebelumnya, $H-1$, $H-2$, ..., $H-k$.
- Ukuran *time series* (k) ditentukan secara coba-coba (*trial-and-error*) karena memang sangat sulit mengetahui berapa jumlah data masukan yang tepat.

Representasi individu GA

$$z = 0,1157 + 0,7315y_1 + 0,3995y_2$$



a_0	a_1	a_2
0,1157	0,7315	0,3995

Grammatical Evolution (GE)

- Untuk masalah prediksi, GE adalah algoritma EAs yang akan menghasilkan model terbaik.
- Dengan representasi kromosom yang berupa fungsi atau program, GE bisa melakukan pencarian model prediksi yang lebih bervariasi
- Dengan membangun *grammar* yang luas dalam notasi *Backus Naur Form* (BNF), GE bisa melakukan pencarian untuk sangat banyak kemungkinan model prediksi, baik linier maupun non-linier.

$N = \{\text{expr}, \text{op}, \text{pre_op}\}$

$T = \{\text{Sin}, \text{Cos}, \text{Tan}, \text{Log}, +, -, /, *, y1, y2, y3, y4, 0,5, 1, 1,5, 2, ()\}$

$S = \langle \text{expr} \rangle$

P dapat direpresentasikan sebagai:

- (1) $\langle \text{expr} \rangle ::= \langle \text{expr} \rangle \langle \text{op} \rangle \langle \text{expr} \rangle$ (A)
| $(\langle \text{expr} \rangle \langle \text{op} \rangle \langle \text{expr} \rangle)$ (B)
| $\langle \text{pre_op} \rangle (\langle \text{expr} \rangle)$ (C)
| $\langle \text{var} \rangle$ (D)

- (2) $\langle \text{op} \rangle ::= +$ (A)
| $-$ (B)
| $/$ (C)
| $*$ (D)

- (3) $\langle \text{pre_op} \rangle ::= \text{Sin}$ (A)
| Cos (B)
| Tan (C)
| Log (D)

- (4) $\langle \text{var} \rangle ::= y1$ (A)
| $y2$ (B)
| $y3$ (C)
| $y4$ (D)
| $0,5$ (E)
| 1 (F)
| $1,5$ (G)
| 2 (H)

Contoh model prediksi GE

$$z = 0,5 + 2y_1 - y_2 + (0,5 * y_3)$$

$$z = 1,5 + \sin(y_1) - \log(y_2)$$

$$z = \frac{0,5 + y_1}{(y_2 + (0,5 * y_3) - \cos(y_4))}$$

Fungsi *Fitness*

$$f = \frac{1}{(K + b)}$$

dimana b merupakan suatu bilangan yang dianggap sangat kecil untuk menghindari pembagian dengan 0, sedangkan K adalah rata-rata kesalahan prediksi untuk semua data penjualan.

Fungsi *Fitness*

Kesalahan prediksi merupakan harga mutlak dari selisih hasil prediksi menggunakan model tersebut (z) dengan data penjualan yang sebenarnya (z^*). Dengan demikian, rata-rata kesalahan prediksi dapat dituliskan sebagai

$$K = \frac{1}{N} \sum_{i=1}^N |z_i - z_i^*|$$

dimana N adalah jumlah semua data prediksi.

Intrusion Detection *Systems* (IDS)

- IDS adalah usaha mengidentifikasi adanya penyusup yang memasuki sistem tanpa otorisasi atau seorang *user* yang sah tetapi menyalahgunakan *privelege* sumber daya sistem.
- Meskipun teknologi ini belum sempurna dan masih memerlukan perbaikan, namun saat ini *intrusion detection* memegang peranan cukup penting pada keseluruhan arsitektur keamanan sebuah sistem jaringan komputer

Jenis Serangan

- ***Probe/scan***

- Usaha-usaha yang tidak lazim untuk memperoleh akses ke dalam suatu sistem atau untuk menemukan informasi tentang sistem tersebut.
- Kegiatan *probe* dalam jumlah besar dengan menggunakan *tools* secara otomatis biasa disebut *scan*.

- ***Denial of Service (DoS)***

- Usaha yang dilakukan untuk membuat sumber daya jaringan maupun komputer tidak bekerja, sehingga tidak mampu memberikan layanan.

- ***Penetration***

- Merupakan sebuah usaha untuk mengubah data, *privelege*, atau sumber daya pada sistem. Beberapa jenis gangguannya, antara lain :
- *User to Root (U2R)*: *user* lokal pada suatu *host* memperoleh hak admin.
- *Remote to user (R2L)*: pengakses luar dapat *account* lokal di *host* target.

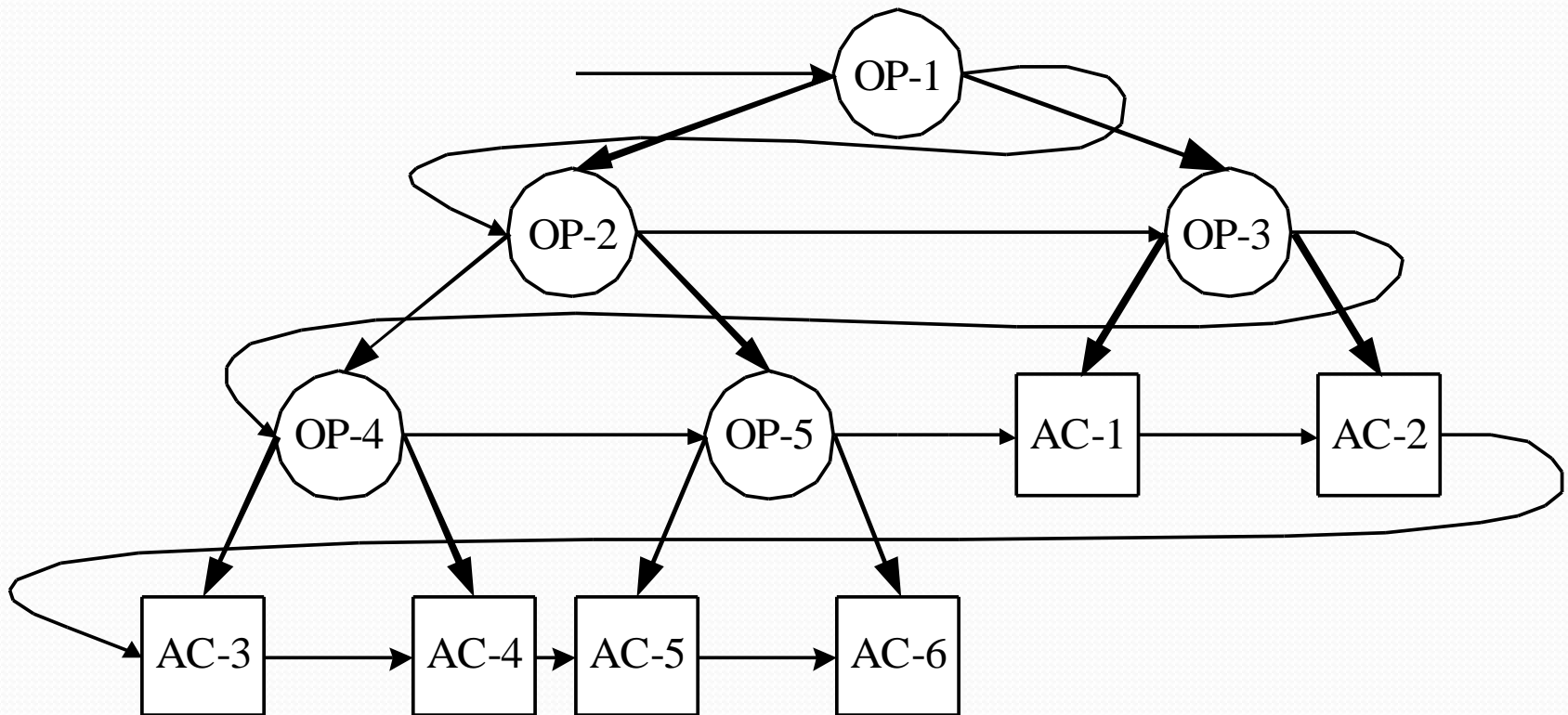
No.	Nama	Keterangan
1	Service	<i>Service</i> yang digunakan pada jaringan.
2	Src_bytes	Jumlah data yang berasal dari <i>node</i> sumber ke tujuan.
3	Dst_bytes	Jumlah data yang berasal dari <i>node</i> tujuan ke sumber.
4	Logged_in	Menunjukkan status <i>login user</i> , jika <i>login</i> sukses maka nilainya TRUE.
5	Count	Jumlah koneksi yang tersambung ke <i>host</i> yang sama dalam 2 detik terakhir.
6	Srv_count	Jumlah koneksi yang tersambung dengan <i>service</i> yang sama dalam 2 detik terakhir.
7	Serror_rate	Prosentase dari koneksi yang memiliki 'SYN' <i>error</i> pada koneksi ke <i>host</i> yang sama.
8	Srv_rerror_rate	Prosentase dari koneksi yang memiliki 'REJ' <i>error</i> pada koneksi dengan <i>service</i> yang sama.
9	Srv_diff_host_rate	Prosentase dari koneksi tersambung ke <i>host</i> yang berbeda.
10	Dst_host_count	Jumlah koneksi yang tersambung ke <i>node</i> tujuan.
11	Dst_host_srv_count	Jumlah koneksi yang tersambung ke <i>node</i> tujuan dengan <i>service</i> yang sama dalam 2 detik terakhir.
12	Dst_host_diff_srv_rate	Jumlah koneksi yang tersambung ke <i>node</i> tujuan dengan <i>service</i> yang berbeda.
13	Class	Kategori data, yaitu normal atau abnormal.

Struktur Linier Kromosom Rule Classifier

Gen ₁				..	Gen _n				Gen _{n+1}		
ac ₁		op ₁		..	ac _n		op _n		ac _{n+1}		*
var ₁	s ₁	o ₁	prec ₁	..	var _n	s _n	o _n	prec _n	var _{n+1}	s _{n+1}	*

- $\langle ac_i \rangle$: *Atomic condition*, terdiri dari: var_i sebagai variabel, dalam hal ini merupakan atribut pada data; s_i sebagai set, yang merupakan label *fuzzy set* untuk nilai suatu data.
- $\langle op_i \rangle$: *Fuzzy operator*, terdiri dari o_i, sebagai operator, yaitu AND dan OR; dan prec_i, sebagai *precedence* atau tingkat prioritas.

Struktur Pohon Kromosom Rule Classifier



Referensi

- [SUYo8] Suyanto, 2008, “Soft Computing: Membangun Mesin Ber-IQ Tinggi”, Informatika, Bandung Indonesia. ISBN: 978-979-1153-49-2.
- [TETo1] Tettamanzi A., Tomassini M., ”Soft Computing”. Springer-Verlag Berlin Heidelberg, 2001. Printed in Germany.