

# Pratikum Manajemen Pemrosesan Data

VI035105



Oleh:

**Ali Ridho Barakbah**

**NIP. 197308162001121001**

**Ira Prasetyaningrum**

**NIP. 198005292008122005**

**Program Studi Teknik Informatika  
Departemen Teknik Informatika dan Komputer  
Politeknik Elektronika Negeri Surabaya  
2012**

# PRAKTIKUM 6

---

## IDENTIFIKASI DATA

---

### A. TUJUAN PEMBELAJARAN

1. Memahami dan mampu Analisa deret waktu.
2. Memahami dan mampu Mencari trend
3. Memahami dan mampu Mencari season
4. Memahami dan mampu Mencari varian

### B. DASAR TEORI

Suatu deret berkala merupakan suatu himpunan observasi dimana variabel yang digunakan diukur dalam urutan periode waktu, misalnya tahunan, bulanan, triwulanan, dan sebagainya. Tujuan dari metode deret berkala adalah untuk menemukan pola data secara historis dan mengekstrapolasikan pola tersebut untuk masa yang akan datang. Peramalan didasarkan pada nilai variabel yang telah lalu dan atau peramalan kesalahan masa lalu.

Peramalan dapat diklasifikasikan berdasarkan:

- a. Kurun waktu yang dapat dicapai oleh ramalan:
  - Jangka panjang
  - Jangka menengah
  - Jangka pendek
- b. Pendekatannya:

- Qalitative atau Judgemental
- Quantitative atau Statistical:
  - Projective : **analisis time series**
  - Causal : analisis regresi

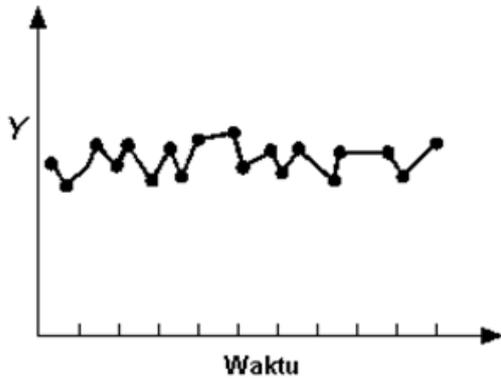
Syarat Peramalan Kuantitatif:

- Tersedia informasi tentang masa lalu
- Informasi tersebut dapat dikuantitatifkan
- Dapat diasumsikan bahwa beberapa aspek pola masa lalu akan terus berlanjut di masa mendatang

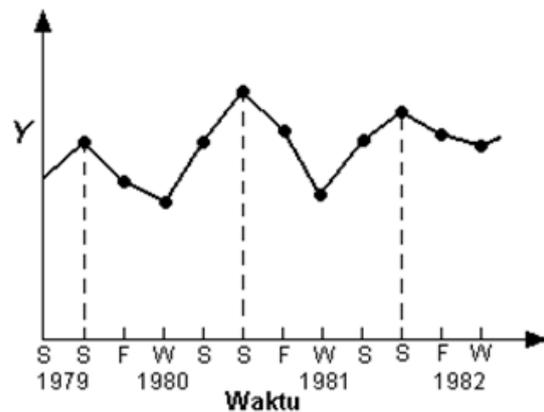
### **Komponen-komponen Deret Berkala**

Apabila kita mengamati Deret Berkala, maka kita akan memperoleh informasi bahwa ada 4 komponen variasi yang penting, yaitu:

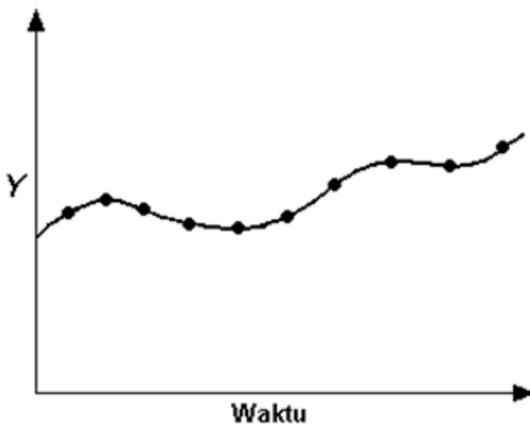
1. **Secular Trend** atau **Trend** (disingkat **T**)
2. **Seasonal Variation** (Variasi Musim) (disingkat **S**)
3. **Cyclical Variation** (Variasi Siklis) (disingkat **C**).
4. **Irregular Variation** (variasi tak beraturan) (disingkat **I**).



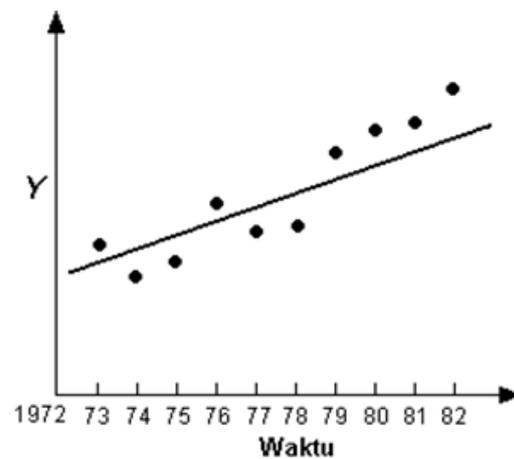
Gambar 6.1. Pola pola Horizontal



Gambar 6.2. Pola Data Musiman



Gambar 6.3. Pola pola Siklis



Gambar 6.4. Pola Data Trend

Pola Data:

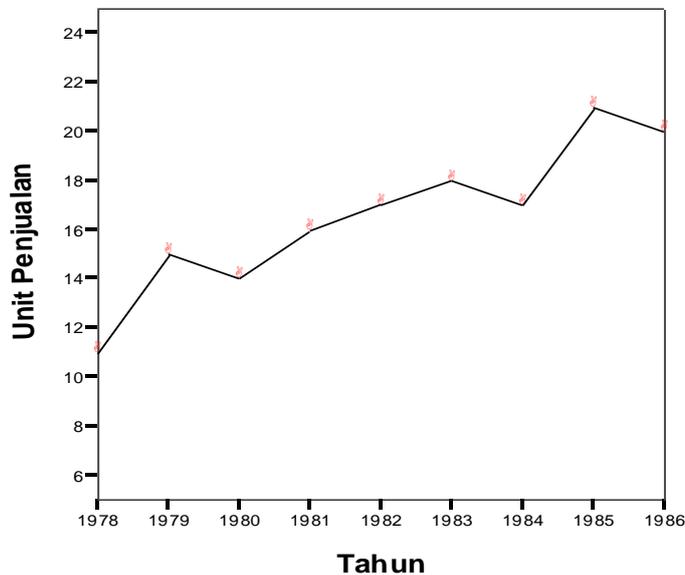
- Pola Horizontal (**H**): nilai data berfluktuasi disekitar nilai rata-rata yang konstan. Mengukur simpangan nilai deret berkala sebenarnya dari yang diharapkan berdasarkan komponen lain. Hal tersebut disebabkan oleh jangka waktu yang pendek (*short-term*) dan faktor yang tidak terantisipasi yang dapat mempengaruhi deret berkala
- Pola Musiman (**S**): terjadi bila duatu data ipengaruhi oleh faktor musiman (kuartal, bulan, minggu, dll). Merepresentasikan pola berulang dengan durasi kurang dari 1 tahun dalam suatu deret berkala. Pola durasi dapat berupa jam atau waktu yang lebih pendek.

- Pola Siklis (C): terjadi bila data dipengaruhi oleh fluktuasi ekonomi jangka panjang yang berhubungan dengan siklus bisnis. Merepresentasikan rangkaian titik-titik dengan pola siklis (pergerakan secara siklis/naik-turun) di atas atau di bawah garis tren dalam kurung waktu satu tahun.
- Pola Trend (T): terjadi bila terdapat kenaikan atau penurunan sekuler jangka panjang dalam data. Merepresentasikan suatu perubahan dari waktu ke waktu (cenderung naik atau turun). Tren biasanya merupakan hasil perubahan dalam populasi/penduduk, faktor demografi, teknologi, dan atau minat konsumen.

Tabel 6.1 Contoh Perhitungan Angka Indeks Penjualan Kendaraan Bermotor Tahun 1978 - 1986

| Tahun | Unit Penjualan |
|-------|----------------|
| 1978  | 11             |
| 1979  | 15             |
| 1980  | 14             |
| 1981  | 16             |
| 1982  | 17             |
| 1983  | 18             |
| 1984  | 17             |
| 1985  | 21             |
| 1986  | 20             |

**Grafik Deret Berkala Penjualan Kendaraan Bermotor**



**Gambar 6.5** Grafik Deret Berkala penjualan kendaraan bermotor

**Secular Trend atau Disingkat Trend → (T)**

**Secular Trend (Trend)** adalah gerak naik, atau turun atau tetap/konstan dalam jangka panjang. Menurut gerakannya dibedakan menjadi tiga, yaitu:

1. Trend naik (*Upward trend*) → trend biaya hidup
2. Trend tetap (*Constant trend*) → trend kapasitas perguruan tinggi
3. Trend turun (*Downward trend*) → trend pekerja di sektor pertanian

**Seasonal Variation (Variasi Musim) → (S)**

**Seasonal variation** (gerak atau variasi musim) adalah gerak naik atau turun *secara periodik* dalam jangka waktu **1 (satu) tahun**.

- Berulang setiap tahun → penjualan pakaian melonjak menjelang hari Lebaran.
- Naik atau turun secara periodik.
- Biasanya dinyatakan dalam persentase. Nilai persentase ini juga disebut dengan istilah *Seasonal Index*. Sebagai contoh, Seasonal Index penjualan pakaian menjelang Lebaran 175% berarti volume penjualan 75% di atas keadaan normal.

**Contoh:**

Tabel 6.2 Data percobaan 1

| Kuartal  | Penjualan Kuartalan |      |      | Spesifikasi Gerak Musim (%) |      |      | Pola Gerak Musim (%) |
|----------|---------------------|------|------|-----------------------------|------|------|----------------------|
|          | 1980                | 1981 | 1982 | 1980                        | 1981 | 1982 |                      |
| I        | 60                  | 50   | 55   | 77                          | 65   | 71   | 71                   |
| II       | 80                  | 90   | 85   | 103                         | 116  | 110  | 110                  |
| III      | 105                 | 100  | 95   | 135                         | 129  | 123  | 129                  |
| IV       | 65                  | 70   | 75   | 84                          | 90   | 97   | 90                   |
| Rerata = | 77.5                | 77.5 | 77.5 |                             |      |      | 100                  |

Spesifikasi Gerak Musim (%) = **(penjualan/rerata) \* 100%**

Contoh →  $(60/77.5)*100\% = 77\%$

$(80/77.5)*100\% = 103\%$

Pola Gerak Musim = **(Spesifikasi 1980 + 1981 + 1982)/3**

Contoh →  $(77+65+71)/3 = 71\%$

$$(103+116+110)/3 = 110\%$$

### **Cyclical Variation (Gerak Siklis) → (C)**

Disebut sebagai **Gerak Siklis** atau **Business Cycle**.

#### **Definisi:**

Gerak Siklis adalah gerak naik atau turun secara periodik dalam jangka panjang, 5 tahun, 10 tahun, 15 tahun, 20 tahun, 25 tahun atau lebih.

#### **Fakta-fakta:**

- Kegiatan ekonomi maupun perusahaan dapat berkembang atau menurun secara periodik dalam jangka lebih dari 1 tahun.
- Tendensi timbulnya Gerak Siklis lebih banyak diakibatkan oleh kegiatan perusahaan, misalnya penjualan mobil, pembangunan gedung, perkembangan tingkat harga, dsb.
- Menurut Gottfried Haberler dalam bukunya **Prosperity and Depression** periode **Business Cycles** dapat dibedakan menjadi 4 bagian, yaitu:
  - Masa kemakmuran (Prosperity phase)
  - Masa Krisis (Downturn, crisis phase)
  - Masa Kehancuran (Depression phase)
  - Masa Pembangunan Kembali (Upturn, Revival Phase).

### **Irregular Variation (Gerak Tak Beraturan) → (I)**

#### **Definisi:**

**Irregular Variation** adalah gerakan tidak teratur dan sulit diramalkan.

#### **Fakta-fakta:**

- Gerakan ini selalu ada pada deret berkala dan sulit dihilangkan.

- Gerakan ini timbul sebagai akibat adanya peperangan, bencana alam, kelaparan, kekeringan, inflasi dan deflasi.
  - Inflasi: suatu kenaikan umum harga rata-rata barang atau jasa selama waktu tertentu dan akibatnya terjadi penurunan daya beli masyarakat sebanding dengan menurunnya nilai mata uang.
  - Deflasi: Penurunan harga rata-rata secara umum barang dan jasa selama jangka waktu tertentu.

### **C. TUGAS PENDAHULUAN**

Untuk masing-masing pola diatas sebutkan masing-masing pola berikan 3 contoh data yang memiliki pola-pola tersebut.

### **D. PERCOBAAN**

Carilah data jumlah pengunjung perpustakaan dalam satu tahun. Analisa data tersebut termasuk dalam pola mana data tersebut. analisa data dengan 3 waktu : waktu penerimaan mahasiswa baru, pada waktu perkuliahan biasa dan pada waktu menjelang ujian ( UTS ataupun UAS).

### **E. LAPORAN RESMI**

Kumpulkan hasil percobaan di atas , Buatlah analisa terhadap hasil tersebut diatas.

# PRAKTIKUM 7

---

## PENGAMBILAN SAMPLE

## DARI POPULASI

---

### A. TUJUAN PEMBELAJARAN

Memahami dan mampu menentukan Jumlah sample yang dibutuhkan yang dapat mewakili populasi.

### B. DASAR TEORI

Alasan Menggunakan Sampel

1. Mengurangi kerepotan
2. Jika populasinya terlalu besar maka akan ada yang terlewat
3. Dengan penelitian sampel maka akan lebih efisien
4. Seringkali penelitian populasi dapat bersifat merusak
5. Adanya bias dalam pengumpulan data
6. Seringkali tidak mungkin dilakukan penelitian dengan populasi

Permasalahan dalam sampel

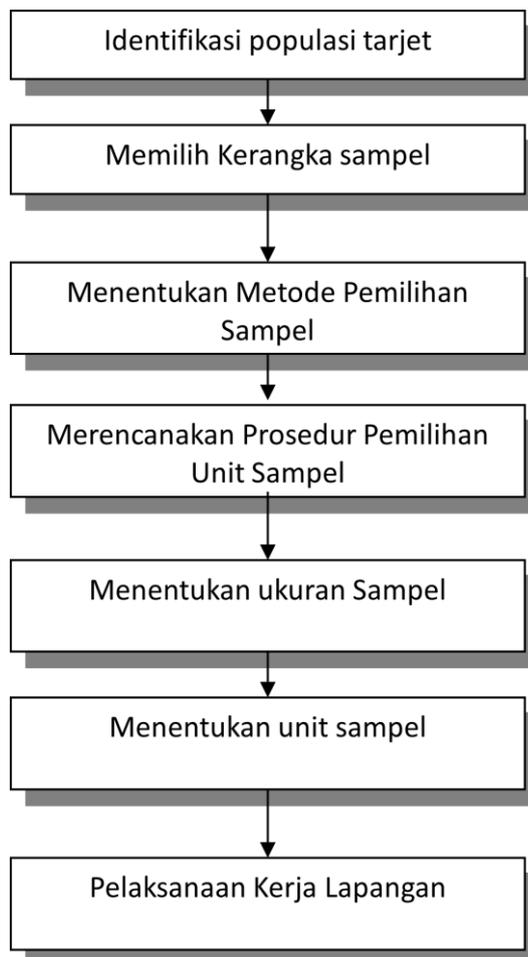
1. Berapa jumlah sampel yang akan diambil
2. Bagaimana teknik pengambilan sampel

Pertimbangan Dalam Menentukan Sampel

1. Seberapa besar keragaman populasi
2. Berapa besar tingkat keyakinan yang kita perlukan

3. Berapa toleransi tingkat kesalahan dapat diterima
4. Apa tujuan penelitian yang akan dilakukan
5. Keterbatasan yang dimiliki oleh peneliti

#### Prosedur penentuan Sample



Gambar 7.1 Prosedur penentuan Sample

#### Menentukan Jumlah Sampel

1. Pendapat Slovin

$$n = \frac{N}{1 + Ne^2}$$

Kita akan meneliti pengaruh upah terhadap semangat kerja pada karyawan PT. Cucak Rowo. Di dalam PT tersebut terdapat 130 orang karyawan. Dengan tingkat kesalahan pengambilan sampel sebesar 5%, berapa jumlah sampel minimal yang harus diambil ?

$$n = \frac{130}{1 + 130(0,05)^2} = 98,11$$

## 2. Interval Penaksiran

Untuk menaksir parameter rata-rata  $\mu$

$$n = \left( \frac{Z_{\alpha/2} \sigma}{e} \right)^2$$

Seorang mahasiswa akan menguji suatu hipotesis yang menyatakan bahwa Indeks Prestasi Mahasiswa Jurusan IT PENS adalah 2,7. dari 30 sampel percobaan dapat diperoleh informasi bahwa standar deviasi indek Prestasi mahasiswa adalah 0,25 Untuk menguji hipotesisi ini berapa jumlah sampel yang diperlukan jika kita menginginkan tingkat keyakinan sebesar 95% dan error estimasi  $\mu$  kurang dari 0,05,?

$$n = \left( \frac{(1,96)(0,25)}{(0,05)} \right)^2 = 96,04$$

- Untuk menaksir parameter proporsi P

$$n = \left( \frac{Z_{\alpha/2}^2 pq}{e^2} \right)$$

Kita akan memperkirakan proporsi mahasiswa yang menggunakan angkutan kota waktu pergi kuliah. Berapa sampel yang diperlukan jika dengan tingkat kepercayaan 95% dan kesalahan yang mungkin terjadi 0,10 ?

$$n = \left( \frac{1,96^2}{4(0,10)^2} \right) = 96,04$$

3. Pendekatan Isac Michel

1. Menentukan sampel untuk menaksir parameter rata-rata  $\mu$

$$n = \frac{NZ^2S^2}{Nd^2 + Z^2S^2}$$

Seorang mahasiswa akan menguji suatu hipotesis yang menyatakan bahwa Indeks Prestasi Mahasiswa Jurusan Manajemen UG yang berjumlah 175 mahasiswa adalah 2,7. Dari 30 sampel percobaan dapat diperoleh informasi bahwa standar deviasi Indeks Prestasi mahasiswa adalah 0,25 Untuk menguji hipotesis ini berapa jumlah sampel yang diperlukan jika kita menginginkan tingkat keyakinan sebesar 95% dan error estimasi  $\mu$  kurang dari 5 persen ?

$$n = \frac{(175)(1,96)^2(0,25)^2}{(175)(0,05)^2 + (1,96)^2(0,25)^2} = 62$$

2. Menentukan sampel untuk menaksir parameter proporsi P

$$n = \frac{NZ^2pq}{Nd^2 + Z^2pq}$$

Kita akan memperkirakan proporsi mahasiswa jurusan manajemen UG yang berjumlah 175 orang. Berdasarkan penelitian pendahuluan diperoleh data proporsi mahasiswa manajemen UG yang menggunakan angkutan kota waktu pergi kuliah adalah 40%. Berapa sampel yang diperlukan jika dengan tingkat kepercayaan 95% dan derajat penyimpangan sebesar 0,10.?

$$n = \frac{(175)(1,96)^2(0,4)(0,6)}{(175)(0,1)^2 + (1,96)^2(0,4)(0,6)} = 60,38$$

**C. TUGAS PENDAHULUAN**

Lakukan pencarian data berapa jumlah seluruh mahasiswa PENS pada saat ini.

**D. PERCOBAAN**

1. Hitunglah sample yang yang diperlukan untuk menganalisa studi kasus tersebut pada pratikum 1.
2. Analisa kelebihan dan kekurangan masing-masing metode perhitungan sample.

**E. LAPORAN RESMI**

Kumpulkan hasil percobaan di atas.dan lakukan analisa.

# PRAKTIKUM 8

---

## PEMROSESAN DATA AWAL

---

### A. TUJUAN PEMBELAJARAN

1. Memahami Pembersihan data
2. Memahami Integrasi data
3. Memahami Transformasi data
4. Memahami Reduksi Data

### B. DASAR TEORI

#### Tugas Utama Pemrosesan Awal Data

- Pembersihan data (data yang kotor)  
Mengisi nilai-nilai yang hilang, menghaluskan noisy data, mengenali atau menghilangkan outlier, dan memecahkan ketidak-konsistenan
- Integrasi data (data heterogen)  
Integrasi banyak database, banyak kubus data, atau banyak file
- Transformasi data (data detail)  
Normalisasi dan agregasi Misal, Salary="\*-10"
- Reduksi data (jumlah data yang besar) .  
Mendapatkan representasi yang direduksi dalam volume tetapi menghasilkan hasil analitis yang sama atau mirip
- Diskritisasi data (kesinambungan atribut)

Bagian dari reduksi data tetapi dengan kepentingan khusus, terutama data numeric

### **Pembersihan Data**

- Kepentingan
  - “Pembersihan data adalah salah satu dari 3 problem terbesar dalam data warehousing”—Ralph Kimball
  - “Pembersihan data adalah problem nomor 1 dalam data warehousing”—DCI survey
- Tugas pembersihan data
  - Mengisi nilai-nilai yang hilang
  - Mengenali outliers dan menghaluskan noisy data
  - Memecahkan redundansi yang disebabkan oleh integrasi data
- Memperbaiki ketak-konsistenan data, US=USA?
  - Menggunakan rujukan eksternal
  - Mendeteksi pelanggaran kendala
    - Misal, kebergantungan fungsional

### **Data Hilang**

- Data tidak selalu tersedia. Misal, banyak tuple atau record tidak memiliki nilai yang tercatat untuk beberapa atribut, seperti customer income dalam data sales. Hilangnya data bisa karena kegagalan pemakaian peralatan, Ketak-konsistenan dengan data tercatat lainnya dan karenanya dihapus.
- Data tidak dimasukkan karena salah pengertian, data tertentu bisa tidak dipandang penting pada saat entry, tidak mencatat history atau tidak mencatat perubahan data

Menangani Data Hilang

- Mengabaikan tuple atau record: mudah tetapi tidak efektif, dan merupakan metoda terakhir. Biasanya dilakukan saat label kelas hilang, Tidak efektif bila persentasi dari nilai-nilai yang hilang per atribut sungguh-sungguh bervariasi.
- Mengisi nilai-nilai yang hilang secara manual: Paling baik, Membosankan, Paling mahal biayanya, dan Tak mungkin dilakukan dalam banyak hal.
- Mengisi nilai-nilai yang hilang secara otomatis, menggunakan: Suatu konstanta global: misal, “unknown”, “Null”, atau suatu kelas baru?! . Suatu pola yang memuat “unknown” atau “Null” adalah buruk. Gunakan rata-rata atribut, Pengempisan data ke mean/median, Rata-rata atribut untuk seluruh sampel yang masuk kedalam kelas yang sama
- Lebih cerdas, dan suatu metoda yang baik. Nilai yang paling mungkin: berbasis inferensi seperti regresi, rumus bayesian, atau pohon keputusan. Klasifikasi untuk mendapatkan nilai yang paling mungkin. Suatu metoda yang baik dengan beberapa overhead
- Menggunakan suatu nilai untuk mengisi nilai yang hilang bisa membiaskan data, nilai bisa salah .
- Nilai yang paling mungkin adalah yang terbaik
- Gunakan informasi yang paling banyak dari data yang ada untuk memprediksi

### **Integrasi Data**

- Integrasi data:  
Mengkombinasikan data dari banyak sumber kedalam suatu simpanan terpadu
- Integrasi skema  
Mengintegrasikan metadata dari sumber-sumber berbeda
- Problem identifikasi entitas: mengenali entitas dunia nyata dari banyak sumber-sumber data, misal  $A.cust-id \equiv B.cust-\#$
- Pendeteksian dan pemecahan konflik nilai data

Untuk entitas dunia nyata yang sama, nilai-nilai atribut dari sumber-sumber berbeda adalah berbeda. Alasan yang mungkin: representasi berbeda, skala berbeda, misal berat bisa dalam pound atau kilogram.

**Contoh Kasus 1:**

- Problem: integrasi skema heterogen
- Nama-nama atribut berbeda

Tabel 8.1 Kasus 1

| cid | name  | byear |
|-----|-------|-------|
| 1   | Jones | 1960  |
| 2   | Smith | 1974  |
| 3   | Smith | 1950  |

| Customer-ID | state |
|-------------|-------|
| 1           | NY    |
| 2           | CA    |
| 3           | NY    |

- Unit berbeda: Sales dalam \$, sales dalam Yen, sales dalam DM



**Gambar 8.1 Kasus 1**

**Contoh Kasus 2:**

- Problem: integrasi skema heterogen
- Skala berbeda: Sales dalam dollar versus sales dalam sen dollar



**Gambar 8.1 Kasus 2**

Atribut turunan: Annual salary versus monthly salary

Tabel 8.2 Kasus 2

| cid | monthlySalary |
|-----|---------------|
| 1   | 5000          |
| 2   | 2400          |
| 3   | 3000          |

| cid | Salary  |
|-----|---------|
| 6   | 50,000  |
| 7   | 100,000 |
| 8   | 40,000  |

**Contoh Kasus 3:**

- Problem: ketak-konsistenan karena redundansi
- Customer dengan customer-id 150 punya 3 anak dalam relation1 dan 4 anak dalam relation2

Tabel 8.3 Kasus 3

| cid | numChildren |
|-----|-------------|
| 1   | 3           |

| cid | numChildren |
|-----|-------------|
| 1   | 4           |

| cid | monthlySalary |
|-----|---------------|
| 1   | 5000          |
| 2   | 6000          |

| cid | Salary |
|-----|--------|
| 1   | 60,000 |
| 2   | 80,000 |

Komputasi annual salary dari monthly salary dalam relation1 tak cocok dengan atribut “annual-salary” dalam relation2

**PenangananRedundansi**

**Dalam Integrasi Data**

- Data redundan sering terjadi saat integrasi dari banyak database
  - Atribut yang sama bisa memiliki nama berbeda dalam database berbeda
  - Atribut yang satu bisa merupakan suatu atribut “turunan” dalam tabel lainnya, misal, annual revenue
- Data redundan mungkin bisa dideteksi dengan analisis korelasi

- Integrasi data hati-hati dari banyak sumber bisa membantu mengurangi/mencegah redundansi dan ketak-konsistenan dan memperbaiki kecepatan dan kualitas mining.
- Suatu atribut adalah redundan jika atribut tersebut bisa diperoleh dari atribut lainnya

- Analisis korelasi

$$r_{A,B} = \frac{\sum (A - \bar{A})(B - \bar{B})}{(n-1)\sigma_A\sigma_B}$$

- Rata-rata A adalah

$$\bar{A} = \frac{\sum A}{n}$$

- Deviasi standard A adalah

$$\sigma_A = \sqrt{\frac{\sum (A - \bar{A})^2}{n-1}}$$

- $R_{A,B} = 0$ : A dan B saling bebas
- $R_{A,B} > 0$ : A dan B berkorelasi positif  $A \uparrow \leftrightarrow B \uparrow$
- $R_{A,B} < 0$ : A dan B berkorelasi negatif  $A \downarrow \leftrightarrow B \uparrow$

### Transformasi Data

- Penghalusan: menghilangkan noise dari data
- Agregasi: ringkasan, konstruksi kubus data
- Generalisasi: konsep hierarchy climbing
- Normalisasi: diskalakan agar jatuh didalam suatu range kecil yang tertentu. Normalisasi min-max, Normalisasi z-score, Normalisasi dengan penskalaan desimal

- Konstruksi atribut/fitur, Atribut-atribut baru dibangun dari atribut-atribut yang ada.

### **Reduksi Data**

- Suatu data warehouse bisa menyimpan terabytes data. Analisis/menambang data kompleks bisa membutuhkan waktu sangat lama untuk dijalankan pada data set komplit (tak efisien)
- Reduksi data , mengurangi ukuran data set tetapi menghasilkan hasil analitis yang sama (hampir sama)
- Strategi reduksi data
  - Agregasi kubus data
  - Reduksi dimensionalitas— menghilangkan atribut tak penting
  - Kompresi data
  - Reduksi Numerosity reduction— mencocokkan data kedalam model
  - Diskritisasi dan pembuatan konsep hierarki

### **C. TUGAS PENDAHULUAN**

Carilah Mengapa data tidak selalu konsisten, tidak lengkap dan terdapat noisy .

### **D. PERCOBAAN**

Pada data jumlah pengunjung perpustakaan dalam satu tahun. Analisa data tersebut apakah data tersebut konsisten, lengkap dan terdapat noisy.

### **E. LAPORAN RESMI**

Kumpulkan hasil percobaan di atas , tambahkan dalam laporan resmi. Beserta hasil uji normalisasi data menggunakan spss.

# PRAKTIKUM 9

---

## IDENTIFIKASI NOISY DATA

---

### A. TUJUAN PEMBELAJARAN

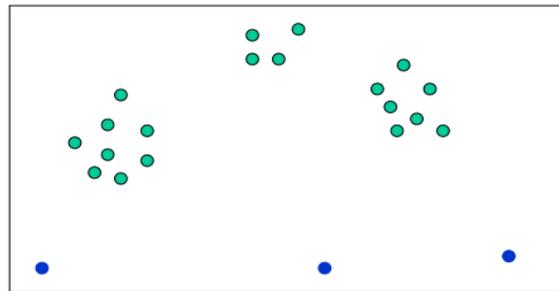
- Memahami Metoda Binning
- Memahami Clustering
- Memahami Inspeksi kombinasi manusia dengan komputer

### B. DASAR TEORI

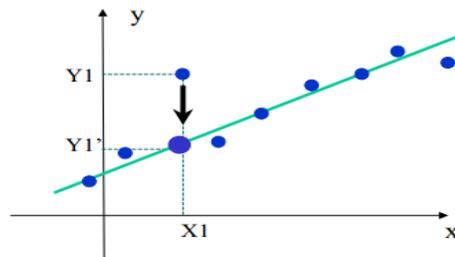
#### Noisy Data

- Noise: error acak atau variansi dalam suatu variable terukur
- Nilai-nilai atribut tak benar mungkin karena
  - Kegagalan instrumen pengumpulan data
  - Problem pemasukan data
  - Problem transmisi data
  - Keterbatasan teknologi
  - Ketak-konsistenan dalam konvensi penamaan
- Problem data lainnya yang memerlukan pembersihan data
  - Duplikasi record
  - Data tak lengkap
  - Data tidak konsisten

## Noisy Data: Menghilangkan Outlier



## Noisy Data: Penghalusan



Menangani Noisy Data?

- Metoda Binning:
  - Pertama urutkan data dan partisi kedalam (kedalaman yang sama) bin-bin
  - Kemudian noisy data itu bisa dihaluskan dengan rata-rata bin, median bin, atau batas bin .
- Clustering
  - Medeteksi dan membuang outliers
- Inspeksi kombinasi komputer dan manusia
  - Mendeteksi nilai-nilai yang mencurigakan dan memeriksa dengan manusia(misal, berurusan dengan outlier yang mungkin)

- Regresi
  - Menghaluskan dengan memasukkan data kedalam fungsi regresi

**Metoda Binning:**

Diskritisasi Sederhana

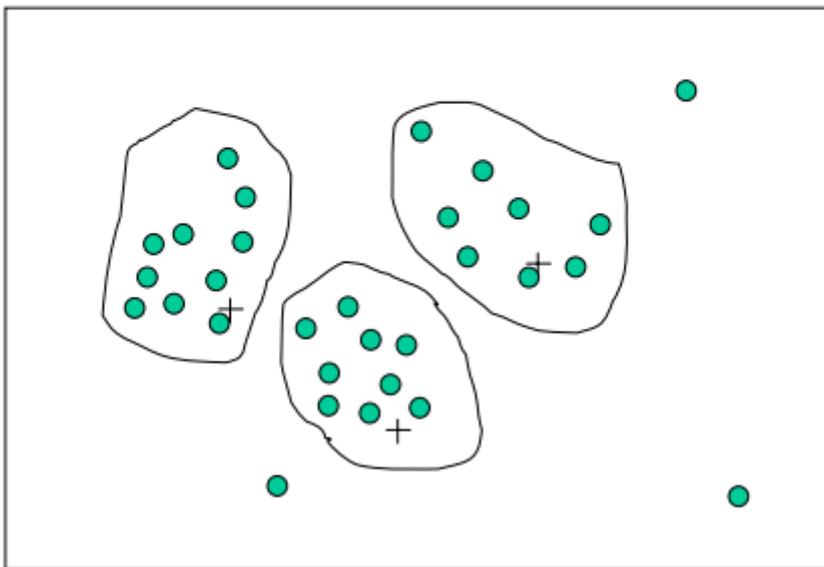
- Partisi lebar yang sama (jarak):
  - Membagi range kedalam N interval dengan ukuran yang sama: grid seragam
  - Jika A dan B masing-masing adalah nilai terendah dan tertinggi dari atribut, lebar interval akan menjadi :  $W = (B - A) / N$ .
  - Kebanyakan langsung, tetapi outlier mendominasi presentasi
  - Data Outlier dan menyimpang tidak ditangani dengan baik.
- Partisi kedalaman sama (frekuensi):
  - Membagi range kedalam N interval, masing-masing memuat jumlah sampel yang hampir sama
  - Penskalaan data yang baik
  - Penanganan atribut yang bersifat kategori bisa rumit.

Metoda Metoda Binning Untuk Penghalusan Data

- Data terurut untuk harga (dalam dollar): 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34
- Partisi kedalam bin dengan kedalaman yang sama (misal, dalam bin-3):
  - Bin 1: 4, 8, 9, 15
  - Bin 2: 21, 21, 24, 25
  - Bin 3: 26, 28, 29, 34
- Haluskan dengan rata-rata bins:
  - Bin 1: 9, 9, 9, 9
  - Bin 2: 23, 23, 23, 23
  - Bin 3: 29, 29, 29, 29
- Penghalusan dengan batas bin:

- Bin 1: 4, 4, 4, 15
- Bin 2: 21, 21, 25, 25
- Bin 3: 26, 26, 26, 34

## Analisis Cluster



### Clustering

Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan(kemiripan).Clustering beda dengan group, kalau group berarti kelompok yang sama,kondisinya kalau tidak ya pasti bukan kelompoknya.Tetapi kalau cluster tidak harus sama akan tetapi pengelompokannya berdasarkan pada kedekatan dari suatu karakteristik sample yang ada, salah satunya dengan menggunakan rumus jarak euclidean.Aplikasinya cluster ini sangat banyak, karena hamper dalam mengidentifikasi permasalahan atau pengambilan keputusan selalu tidak sama persis akan tetapi cenderung memiliki kemiripan saja.

- Mempartisi data set kedalam cluster-cluster, dan bisa hanya menyimpan representasi cluster
- Bisa sangat efektif jika data di-cluster tetapi tidak jika data “dirusak”
- Bisa memiliki clustering hierarki dan bisa disimpan didalam struktur pohon indeks multi-dimensional
- Ada banyak pilihan dari definisi clustering dan algoritma clustering.

**Prinsip dasar :**

- Similarity Measures (ukuran kedekatan)
  - Distances dan Similarity Coeficients untuk beberapa sepasang dari item
- Eclidean Distance:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

Atau :

$$d(x, y) = \left[ \sum_{i=1}^p |x_i - y_i|^2 \right]^{1/2}$$

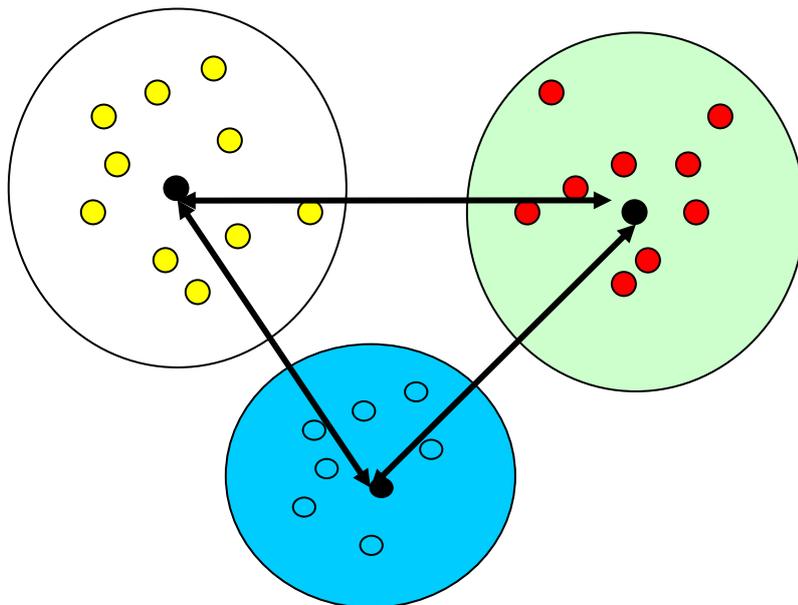
**4. Metode Clustering pada dasarnya mengoptimalkan pusat cluster(centroid) atau mengoptimalkan lebar antar cluster.**

Macam-macam metode clustering :

- Berbasis Metode Statistikk
  - a. Hirarchical clustering method : pada kasus untuk jumlah kelompok belum ditentukan terlebih dulu, contoh data-data hasil survey kuisisionerMacam-metode jenis ini: Single Linkage, Complete Linkage, Average Linkage dll.

b. Non Hierarchical clustering method: Jumlah kelompok telah ditentukan terlebih dulu. Metode yang digunakan : K-Means.

- Berbasis Fuzzy : Fuzzy C-Means
- Berbasis Neural Network : Kohonen SOM, LVQ
- Metode lain untuk optimasi centroid atau lebar cluster : Genetik Algoritma (GA)



Gambar : Salah contoh 3 cluster yang terbentuk

### Pembahasan Metode Cluster

Pertama akan kita bahas dulu metode cluster secara statistic untuk non hirachical method yaitu: K-Means Clustering

Algoritma:

1. Partisi item menjadi K initial cluster

2. Lakukan proses perhitungan dari daftar item, tandai item untuk kelompok yang mana berdasarkan pusat(mean) yang terdekat (dengan menggunakan distance dapat digunakan Euclidean distance). Hitung kembali pusat centroid untuk item baru yang diterima pada cluster tersebut dari cluster yang kehilangan item.
3. Ulangi step 2 hingga tidak ada lagi tempat yang akan ditandai sebagai cluster baru.

Contoh :

Diketahui data sebagai berikut:

| Item | Observasi |    |
|------|-----------|----|
|      | X1        | X2 |
| A    | 5         | 3  |
| B    | -1        | 1  |
| C    | 1         | -2 |
| D    | -3        | -2 |

Ditanya: Lakukan cluster menjadi 2 kelompok (K=2).

- Langkah pertama:

| Cluster | Koordinate dari Centroid  |                            |
|---------|---------------------------|----------------------------|
|         | $\bar{X}_1$               | $\bar{X}_2$                |
| (AB)    | $\frac{5 + (-1)}{2} = 2$  | $\frac{3 + 1}{2} = 2$      |
| (CD)    | $\frac{1 + (-3)}{2} = -1$ | $\frac{-2 + (-2)}{2} = -2$ |

- Langkah kedua :

Lakukan perhitungan jarak dengan eclidean dari masing-masing item dari centroid(pusat) cluster dan tandai kembali setiap item berdasarkan kedekatan group. Jika item bergerak dari initial configuration, Centroid(pusat/means) cluster harus diupdate sebelum diproses. Kita hitung kwadrat jarak(squared distance) sbb:

$$d^2(A, (AB)) = (5 - 2)^2 + (3 - 2)^2 = 10$$

$$d^2(A, (CD)) = (5 + 1)^2 + (3 + 2)^2 = 61$$

Sejak A dekat pada cluster (AB) dibandingkan pada cluster (CD), maka tidak perlu ditandai. Lanjutkan perhitungan :

$$d^2(B, (AB)) = (-1 + 2)^2 + (1 - 2)^2 = 10$$

$$d^2(B, (CD)) = (-1 + 1)^2 + (1 + 2)^2 = 9$$

Sehingga B akan ditandai kembali menjadi anggota baru pada cluster (CD), sehingga membentuk cluster baru (BCD) maka koordinat dari pusat cluster terupdate sebagai berikut :

| Cluster | Koordinate dari Centroid |             |
|---------|--------------------------|-------------|
|         | $\bar{X}_1$              | $\bar{X}_2$ |
| A       | 5                        | 3           |
| (BCD)   | -1                       | -1          |

Selanjutnya lakukan cek untuk setiap item untuk ditandai kembali. Perhitungan kwadrat jarak(squared distances) dibagikan sbb:

| Cluster | Koordinate dari Centroid |    |    |    |
|---------|--------------------------|----|----|----|
|         | A                        | B  | C  | D  |
| A       | 0                        | 40 | 41 | 89 |
| (BCD)   | 52                       | 4  | 5  | 5  |

Kita lihat setiap item yang baru telah ditandai untuk cluster berdasarkan centroid(pusat) terdekat maka proses telah dihentikan. Sehingga dengan K=2 cluster maka terbentuk cluster sebagai berikut : A dan (BCD).

### **Inspeksi Komputer dan Manusia Penghalusan**

- Inspeksi kombinasi komputer dan manusia
  - Suatu ambang yang diberikan user
  - Komputer mendeteksi seluruh potensi outlier yang dikaitkan dengan ambang
  - Manusia menentukan outlier sesungguhnya

### **C. TUGAS PENDAHULUAN**

Carilah makalah yang menggunakan metode clustering

### **D. PERCOBAAN**

Buatlah Program untuk mengklasterkan data menggunakan satu metode clustering.

### **E. LAPORAN RESMI**

Kumpulkan hasil percobaan di atas , tambahkan dalam laporan resmi.

# PRAKTIKUM 10

---

## NORMALISASI DATA

---

### A. TUJUAN PEMBELAJARAN

- Memahami Metode Min Max
- Memahami Metode Z-Score
- Memahami Metode Decimal scaling

### B. DASAR TEORI

#### Transformasi Data

- Penghalusan: menghilangkan noise dari data
- Agregasi: ringkasan, konstruksi kubus data
- Generalisasi: konsep hierarchy climbing
- Normalisasi: diskalakan agar jatuh didalam suatu range kecil yang tertentu
  - Normalisasi min-max
  - Normalisasi z-score
  - Normalisasi dengan penskalaan desimal
- Konstruksi atribut/fitur
  - Atribut-atribut baru dibangun dari atribut-atribut yang ada

- Normalisasi min-max

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

- Normalisasi z-score (saat Min, Max tak diketahui)

$$v' = \frac{v - \text{mean}_A}{\text{stand\_dev}_A}$$

- Normalisasi dengan penskalaan desimal

$$v' = \frac{v}{10^j} \quad \text{dimana } j \text{ adalah integer terkecil sehingga } \text{Max}(|v'|) < 1$$

#### *Min-max Normalization*

Metode normalisasi ini menghasilkan transformasi *linier* pada data asal. Bila  $\min_A$  dan  $\max_A$  adalah nilai minimum dan maksimum dari sebuah atribut A, *Min-max Normalization* memetakan sebuah nilai  $v$  dari A menjadi  $v'$  dalam *range* nilai minimal dan maksimal yang baru,  $\text{new\_min}_A$  dan  $\text{new\_max}_A$ .

### **C. TUGAS PENDAHULUAN**

Sebutkan keuntungan dan kelemahan dari metode min max, z-score dan penskalaan decimal.

### **D. PERCOBAAN**

Aplikasikan metode min max, z-score dan penskalaan pada percobaan studi kasus pratikum 1

### **E. LAPORAN RESMI**

Buatlah laporan resmi dari percobaan diatas, buatlah program aplikasinya juga.

# PRAKTIKUM 11

---

## KORELASI DATA

---

### A. TUJUAN PEMBELAJARAN

1. Memahami Korelasi linier
2. Memahami Korelasi ganda dan parsial

### B. DASAR TEORI

Analisis korelasi adalah alat yang membahas tentang derajat hubungan antara satu variabel dengan variabel lainnya.

Dua variabel dikatakan berkorelasi apabila perubahan dalam satu variabel diikuti oleh perubahan variabel lain, baik yang searah maupun tidak. Hubungan antara variabel dapat dikelompokkan menjadi tiga jenis :

#### 1) Korelasi Positif

Terjadinya korelasi positif apabila perubahan antara variabel yang satu diikuti oleh variabel lainnya dengan arah yang sama (berbanding lurus). Artinya apabila variabel yang satu meningkat, maka akan diikuti peningkatan variabel lainnya.

#### 2) Korelasi Negatif

Terjadinya korelasi negatif apabila perubahan antara variabel yang satu diikuti oleh variabel lainnya dengan arah yang berlawanan (berbanding terbalik). Artinya apabila variabel yang satu meningkat, maka akan diikuti penurunan variabel lainnya.

#### 3) Korelasi Nihil

Terjadinya korelasi nihil apabila perubahan antara variabel yang satu diikuti oleh variabel lainnya dengan arah yang tidak teratur (acak). Artinya apabila variabel yang satu meningkat, maka akan diikuti penurunan variabel. Artinya apabila variabel yang satu meningkat, kadang diikuti dengan peningkatan pada variabel lain dan kadang diikuti dengan penurunan pada variabel lain.

Berdasarkan hubungan antar variabel yang satu dengan variabel lainnya dinyatakan dengan koefisien korelasi yang disimbolkan dengan “  $r$  “. besarnya koefisien korelasi berkisar antara  $-1$   $r$   $+1$

Untuk mencari korelasi antara variabel  $Y$  terhadap  $X_1$  atau  $r_{y.1,2,\dots,k}$  dapat dicari

$$r_{y.1,2,\dots,k} = \frac{n \sum X_{1i} Y_i - (\sum X_{1i})(\sum Y_i)}{\sqrt{(n \sum X_{1i}^2 - (\sum X_{1i})^2)(n \sum Y_i^2 - (\sum Y_i)^2)}}$$

dengan rumus :

Sedangkan untuk mengetahui korelasi antar variabel bebas dengan tiga buah variabel bebas adalah :

a. Koefisien korelasi antara  $X_1$  dan  $X_2$

$$r_{12} = \frac{n \sum X_{1i} X_{2i} - (\sum X_{1i})(\sum X_{2i})}{\sqrt{(n \sum X_{1i}^2 - (\sum X_{1i})^2)(n \sum X_{2i}^2 - (\sum X_{2i})^2)}}$$

b. Koefisien Korelasi antara  $X_1$  dan  $X_3$

$$r_{13} = \frac{n \sum X_{1i} X_{3i} - (\sum X_{1i})(\sum X_{3i})}{\sqrt{(n \sum X_{1i}^2 - (\sum X_{1i})^2)(n \sum X_{3i}^2 - (\sum X_{3i})^2)}}$$

c. Koefisien Korelasi antara  $X_2$  dan  $X_3$

$$r_{23} = \frac{n \sum X_{2i} X_{3i} - (\sum X_{2i})(\sum X_{3i})}{\sqrt{(n \sum X_{2i}^2 - (\sum X_{2i})^2)(n \sum X_{3i}^2 - (\sum X_{3i})^2)}}$$

Nilai koefisien korelasi adalah  $-1$   $r$   $+1$ . Jika dua variabel berkorelasi negatif maka nilai koefisien korelasinya akan mendekati  $-1$  ; jika dua variabel tidak berkorelasi maka nilai koefisien korelasinya akan mendekati  $0$  ; sedangkan jika dua variabel berkorelasi positif maka nilai koefisien korelasinya akan mendekati  $1$ .

Untuk lebih mengetahui seberapa jauh derajat antara variabel – variabel tersebut, dapat dilihat dalam perumusan berikut :

|                           |  |
|---------------------------|--|
| $-1,00 \leq r \leq -0,80$ | Berarti korelasi kuat secara negatif   |
| $-0,79 \leq r \leq -0,50$ | Berarti korelasi sedang secara negatif |
| $-0,49 \leq r \leq 0,49$  | Berarti korelasi lemah                 |
| $0,50 \leq r \leq 0,79$   | Berarti korelasi sedang secara positif |
| $0,80 \leq r \leq 1,00$   | Berarti korelasi kuat secara positif   |

### **C. TUGAS PENDAHULUAN**

Pelajari cara mengkorelasikan data menggunakan SPSS.

### **D. PERCOBAAN**

Pada Data Pengunjung Perpustakaan, korelasikan antar jumlah pengunjung dengan variable yang menyebabkan pengunjung perpustakaan meningkat.

### **E. LAPORAN RESMI**

Kumpulkan hasil percobaan di atas ,Buatlah laporan resmi dan analisa hasil tersebut, menggunakan SPSS.

# PRAKTIKUM 12

---

## REGRESI

---

### A. TUJUAN PEMBELAJARAN

1. Memahami Regresi linier,
2. Memahami Regresi polinomial

### B. DASAR TEORI

Untuk melakukan peramalan diperlukan metode tertentu dan metode mana yang digunakan tergantung dari data dan informasi yang akan diramal serta tujuan yang hendak dicapai. Dalam prakteknya terdapat berbagai metode peramalan antara lain :

#### 1. Time Series atau Deret Waktu

Analisis time series merupakan hubungan antara variabel yang dicari (dependent) dengan variabel yang mempengaruhinya (independent variable), yang dikaitkan dengan waktu seperti mingguan, bulan, triwulan, catur wulan, semester atau tahun.

Dalam analisis time series yang menjadi variabel yang dicari adalah waktu.

Metode peramalan ini terdiri dari :

- a. Metode Smoting, merupakan jenis peramalan jangka pendek seperti perencanaan persediaan, perencanaan keuangan. Tujuan penggunaan metode ini adalah untuk mengurangi ketidakaturan data masa lampau seperti musiman.
- b. Metode Box Jenkins, merupakan deret waktu dengan menggunakan model matematis dan digunakan untuk peramalan jangka pendek.

- c. Metode proyeksi trend dengan regresi, merupakan metode yang digunakan baik untuk jangka pendek maupun jangka panjang. Metode ini merupakan garis trend untuk persamaan matematis.

## 2. Causal Methods atau sebab akibat

Merupakan metode peramalan yang didasarkan kepada hubungan antara variabel yang diperkirakan dengan variabel lain yang mempengaruhinya tetapi bukan waktu. Dalam prakteknya jenis metode peramalan ini terdiri dari :

- a. Metode regresi dan korelasi, merupakan metode yang digunakan baik untuk jangka panjang maupun jangka pendek dan didasarkan kepada persamaan dengan teknik least squares yang dianalisis secara statis.
- b. Model Input Output, merupakan metode yang digunakan untuk peramalan jangka panjang yang biasa digunakan untuk menyusun trend ekonomi jangka panjang.
- c. Model ekonometri, merupakan peramalan yang digunakan untuk jangka panjang dan jangka pendek.

## METODE REGRESI

Penggunaan metode ini didasarkan kepada variabel yang ada dan yang akan mempengaruhi hasil peramalan. Hal-hal yang perlu diketahui sebelum melakukan peramalan dengan metode regresi adalah mengetahui terlebih dahulu mengetahui kondisi-kondisi seperti :

- a. Adanya informasi masa lalu
- b. Informasi yang ada dapat dibuatkan dalam bentuk data (dikuantifikasikan)
- c. Diasumsikan bahwa pola data yang ada dari data masa lalu akan berkelanjutan dimasa yang akan datang.

Adapun data-data yang ada dilapangan adalah :

- a. Musiman (Seasonal)

- b. Horizontal (Stationary)
- c. Siklus (Cylikal)
- d. Trend

Dalam menyusun ramalan pada dasarnya ada 2 macam analisis yang dapat digunakan yaitu :

1. Analisis deret waktu (Time series), merupakan analisis antarvariabel yang dicari dengan variabel waktu
2. Analisis Cross Section atau sebab akibat (Causal method), merupakan analisis variabel yang dicari dengan variabel bebas atau yang mempengaruhi.

Ada dua pendekatan untuk melakukan peramalan dengan menggunakan analisis deret waktu dengan metode regresi sederhana yaitu :

1. Analisis deret waktu untuk regresi sederhana linier
2. Analisis deret untuk regresi sederhana yang non linier

Untuk menjelaskan hubungan kedua metode ini kita gunakan notasi matematis seperti :

$$Y = F(x)$$

Dimana :

Y = Dependent variable (variabel yang dicari)

X = Independent variable (variabel yang mempengaruhinya)

Notasi regresi sederhana dengan menggunakan regresi linier (garis lurus) dapat digunakan sebagai berikut :

$$Y = a + b x$$

Dimana a dan b adalah merupakan parameter yang harus dicari. Untuk mencari nilai a dapat digunakan dengan menggunakan rumus :

$$a = \bar{Y} - b \bar{x}$$

kemudian nilai b dapat dicari dengan rumus :

$$b = \frac{\sum XY - \bar{X} \bar{Y}}{\sum X^2 - \bar{X} \sum X}$$

### ANALISIS DENGAN REGRESI LINIER CROSS SECTION

Cross section method atau casual method atau sebab akibat merupakan peramalan yang kita lakukan untuk mengukur peramalan dalam suatu periode dengan faktor yang mempengaruhinya bukan waktu.

Penggunaan rumusan yang kita gunakan untuk cross section sama dengan penerapan untuk metode time series, begitu pula dengan hasil peramalannya.

Jadi penjualan = f (x, x, x,.....)

X = harga, mutu pendapatan, promosi dll

Y = a + b x

Dimana x adalah variabel bukan waktu.

### ANALISIS DERET WAKTU DENGAN REGRESI LINIER

Ada 2 pendekatan untuk melakukan peramalan dengan menggunakan analisis deret waktu dengan metode regresi sederhana, yaitu :

1. Analisis deret waktu untuk regresi sederhana linier
2. Analisis deret waktu untuk regresi sederhana yang non linier

Dalam analisis deret waktu yang linier adalah analisis pola hubungan yang dicari dengan satu variabel yang mempengaruhinya : waktu. Sedangkan analisis deret waktu yang non linier, merupakan analisis hubungan antara variabel yang dicari dengan hanya satu (1) yang mempengaruhinya, yaitu variabel waktu.

Untuk menjelaskan hubungan kedua metode ini kita gunakan notasi matematis seperti :

$$Y = F(x)$$

Dimana :

Y = Dependent variable (variabel yang dicari)

X = Independent variable (variabel yang mempengaruhinya)

Notasi regresi sederhana dengan menggunakan regresi linier (garis lurus) dapat digunakan sebagai berikut :

$$Y = a + b X \dots\dots\dots$$

Dimana a dan b adalah merupakan parameter (koefisien regresi) yang harus dicari.

Untuk mencari nilai a dapat digunakan dengan menggunakan rumus :

$$a = \frac{\sum Y}{n} - b \frac{\sum X}{n}$$

atau :

$$a = \bar{Y} - b \bar{X}$$

kemudian nilai b dapat dicari dengan rumus :

$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

atau

$$b = \frac{\sum XY - \bar{X} \sum Y}{\sum X^2 - \bar{X} \sum X}$$

Langkah selanjutnya adalah melakukan pengujian terhasi yang diperoleh dengan :

*Pertama*, uji Test Koefisien Penentu ( $R^2$ ), pengtesan ini untuk mengetahui tepat tidaknya varibel yang mempengaruhi besarnya penjualan yang diramalkan adalah waktu.

*Kedua*, Test Significance (T.Test) atau F test yaitu pengtesan untuk mengetahui apakah benar persamaan regresi itu adalah linier.

***Pengujian R<sup>2</sup>.***

Test rumusan yang digunakan adalah:

$$R^2 = 1 - \frac{\sum e^2}{\sum y^2}$$

Dimana,

$$\sum e^2 = Y^2 - b^2 (\sum x^2)$$

$$\sum y^2 = Y^2 - n (\bar{Y})^2$$

$$\sum x^2 = X^2 - n (\bar{X})^2$$

***Test Significance***

Tujuan test ini menguji dan meneliti apakah regresi yang digunakan dalam menyusun ramalan adalah benar linier, dimana data yang diteliti tepat berada disekitar garis linier.

1. F. Test

Tujuannya adalah untuk mengetahui apakah nilai estimasi dari a dan b dapat bervariasi karena pengaruh sampling/ random.

Persamaan F. Test adalah sebagai berikut :

$$F = \frac{\sum (Y - \hat{Y})^2}{k - 1} \bigg/ \frac{\sum (Y - \bar{Y})^2}{n - k}$$

Dimana :

k = jumlah variabel (dalam regresi sederhana = 2)

n = jumlah tahun

atau kita juga menggunakan rumus sebagai berikut :

$$F = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}}$$

Hasil  $F_{\text{rasio}}$  kemudian kita bandingkan dengan  $F_{\text{tabel}}$  apabila  $F_{\text{rasio}} > F_{\text{tabel}}$ , maka secara statistik koefisien b adalah *significance* berbeda dengan nol (0), sehingga persamaan regresi dapat dilakukan secara benar dengan bentuk persamaan sebagai berikut :  $Y = a + b x$

Demikian pula sebaliknya jika  $F_{\text{rasio}} < F_{\text{tabel}}$

## 2. Persamaan T. Test

Test ini dikenal dengan nama student-t didistribusikan untuk menguji a dan b dengan formula :

$$t_{\text{test a}} = \frac{a}{\tau a} \qquad t_{\text{test b}} =$$

Hasilnya jika diperoleh :

$T_{\text{test}} > T_{\text{tabel}}$  ( $T_{\text{distribusi}}$ ), maka tingkat keyakinan tertentu (R) dapat disimpulkan bahwa nilai koefisien regresi a dan b secara statistic berbeda dari (0) dan demikian pula sebaliknya.

## ANALISA DERET WAKTU DENGAN REGRESI NON LINIER

Analisa deret waktu dengan regresi non linier merupakan regresi bukan garis lurus. Notasi regresi sederhana dengan menggunakan regresi linier (garis lurus) dapat digunakan sebagai berikut :

$$Y = a + b x + c x^2$$

Dimana :Y = Dependent variable (variabel yang dicari)

x = Independent variable (variabel yang mempengaruhinya)

a = b = c = parameter koefisien regresi

Formula umum yang digunakan sebagai berikut :

$$\sum y = n a + b \sum x + c \sum x^2$$

$$\sum xy = a \sum x + b \sum x^2 + c \sum x^3$$

$$\sum xy = a \sum x^2 + b \sum x^3 + c \sum x^4$$

### C. TUGAS PENDAHULUAN

Untuk semua persoalan di bawah ini, desainlah algoritma dan flowchartnya :

### D. PERCOBAAN

Carilah dari regresi Data Berikut,

Dan carilah nilai y jika x=3. Gunakan metode-metode diatas.

| i | xi | yi |
|---|----|----|
|---|----|----|

|   |      |        |
|---|------|--------|
| 1 | 0.15 | 4.4964 |
| 2 | 0.4  | 5.1284 |
| 3 | 0.6  | 5.6931 |
| 4 | 1.01 | 6.2884 |
| 5 | 1.5  | 7.0989 |
| 6 | 2.2  | 7.5507 |
| 7 | 2.4  | 7.5106 |

**E. LAPORAN RESMI**

Kumpulkan hasil percobaan di atas , tambahkan dalam laporan resmi hasil analisa perbandingan dari beberapa metode diatas.

# PRAKTIKUM 13

---

## PERAMALAN DENGAN TIME SERIES\_1

---

### A. TUJUAN PEMBELAJARAN

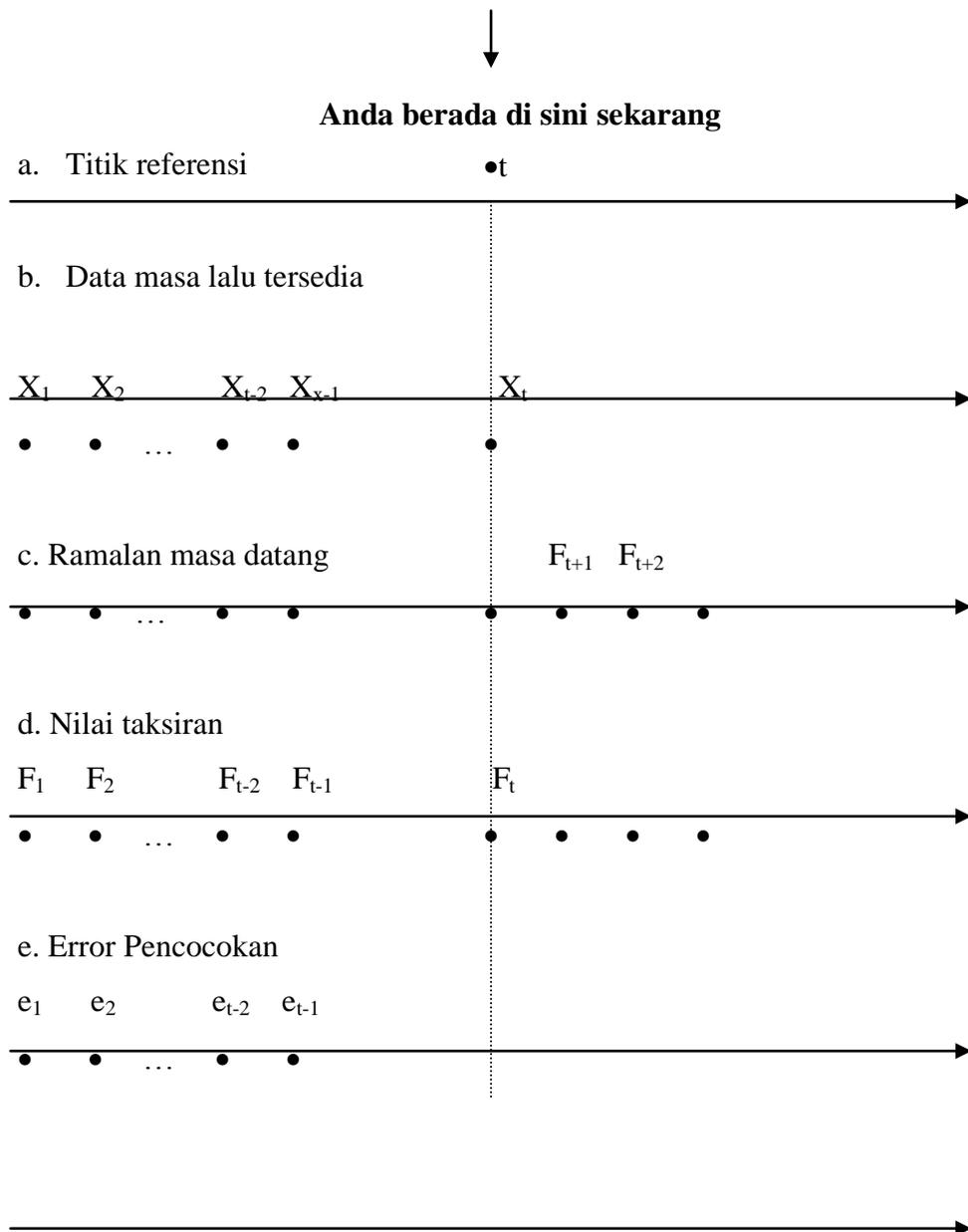
1. Memahami Moving average ,
2. Memahami weight moving average,

### B. DASAR TEORI

*Definisi:*

**Analisis Deret Berkala** (*time series analysis*) adalah suatu metode kuantitatif untuk menentukan pola data masa lampau yang telah dikumpulkan secara teratur menurut urutan waktu kejadian. Pola masa lalu ini dapat digunakan sebagai dasar pertimbangan untuk forecasting di masa yang akan datang.

Skenario Peramalan:



f. Error Peramalan

$e_{t+1}$   $e_{t+2}$

• • •

**METODE PROYEKTIF:**

1. Metode Sederhana:

a. **Naïve Model:**

$$F_{t+1} = X_t$$

b. **Average Change Model:**

$$F_{t+1} = X_t + (X_t - X_{t-1})$$

c. **Average Percent Change Model:**

$$F_{t+1} = X_t \times \frac{X_t}{X_{t-1}}$$

2. Metode Rata-Rata:

a. **Metode Rata-rata Sederhana:**

$$F_{t+1} = \bar{X}$$

b. **Metode Rata-rata Bergerak:**

$$F_{t+1} = \frac{X_1 + X_2 + X_3 + \dots + X_t}{t}$$

$$F_{t+2} = \frac{X_2 + X_3 + \dots + X_t + X_{t+1}}{t}$$

c. **Metode Rata-rata Bergerak Ganda:**

$$S'_t = \frac{X_t + X_{t-1} + X_{t-2} + \dots + X_{t-n+1}}{n}$$

$$S''_t = \frac{S'_t + S'_{t-1} + S'_{t-2} + \dots + S'_{t-n+1}}{n}$$

$$a_t = S'_t + (S'_t - S''_t) = 2S'_t - S''_t$$

$$b_t = \frac{2}{n-1} (S'_t - S''_t)$$

$$F_{t+m} = a_t + b_t m$$

Contoh:

Diketahui Data Bulanan Perjalanan Penumpang Udara selama 5 tahun:

|         | Jan | Peb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Tahun-1 | 171 | 180 | 193 | 181 | 183 | 218 | 230 | 242 | 209 | 191 | 172 | 194 |
| Tahun-2 | 196 | 196 | 236 | 235 | 229 | 243 | 264 | 272 | 237 | 211 | 180 | 201 |
| Tahun-3 | 204 | 188 | 235 | 227 | 234 | 264 | 302 | 293 | 259 | 229 | 203 | 229 |
| Tahun-4 | 242 | 233 | 267 | 269 | 270 | 315 | 364 | 347 | 312 | 274 | 237 | 278 |
| Tahun-5 | 284 | 277 | 317 | 313 | 318 | 374 | 413 | 405 | 355 | 306 | 271 | 306 |

Hasil Peramalan Dengan Minitab, Untuk Berbagai Metode:

### **Moving Average**

Length: 3

Accuracy Measures

MAPE: 11.88

MAD: 31.02

MSD: 1474.06

| Row | Period | Forecast | Lower   | Upper   |
|-----|--------|----------|---------|---------|
| 1   | 61     | 294.333  | 219.082 | 369.585 |
| 2   | 62     | 294.333  | 219.082 | 369.585 |
| 3   | 63     | 294.333  | 219.082 | 369.585 |
| 4   | 64     | 294.333  | 219.082 | 369.585 |

### **Manfaat Analisis Deret Berkala (Time Series Analysis)**

- Membantu mempelajari data masa lampau, sehingga dapat dipelajari faktor-faktor penyebab perubahan untuk pertimbangan perencanaan di masa yang akan datang.
- Untuk membantu dalam peramalan (forecasting).
- Membantu memisahkan faktor-faktor yang dapat mempengaruhi suatu data (khususnya variasi atau gerak musim) → lalu diadakan penyesuaian dengan faktor musim ini.

- Membantu dan mempermudah membandingkan satu rangkaian data dengan rangkaian data yang lain.

### C. TUGAS PENDAHULUAN

Manfaat peramalan untuk kehidupan. Berikan contoh dan jelaskan.

### D. PERCOBAAN

| <u>Bulan</u> | <u>t</u> | <u>d<sub>t</sub></u>            |
|--------------|----------|---------------------------------|
| Jan          | 1        | 90                              |
| Feb          | 2        | 111                             |
| Mar          | 3        | 99                              |
| Apr          | 4        | 89                              |
| Mei          | 5        | 87                              |
| Jun          | 6        | 84                              |
| Jul          | 7        | 104                             |
| Aus          | 8        | 102                             |
| Sep          | 9        | 95                              |
| Okt          | 10       | 114                             |
| Nov          | 11       | 103                             |
| Des          | 12       | <u>113</u>                      |
|              |          | <b><math>\Sigma=1191</math></b> |

Gunakan Metode Moving Average untuk meramal demand untuk tahun ke 13 dan hitung MAD nya(gunakan periode 3,5,7 bulan dan bandingkan MAD nya)

**E. LAPORAN RESMI**

Kumpulkan hasil percobaan di atas , tambahkan dalam laporan resmi . buatlah program aplikasinya.

# PRAKTIKUM 14

---

## PERAMALAN DENGAN TIME SERIES\_2

---

### A. TUJUAN PEMBELAJARAN

Memahami Exponensial smoothing dan metode winter.

### B. DASAR TEORI

#### Metode Pemulusan Eksponensial

##### a. Metode Pemulusan Eksponensial Tunggal

**Dasar:**

$$F_{t+1} = \frac{X_1 + X_2 + X_3 + \dots + X_t}{t}$$
$$F_{t+2} = \frac{X_2 + X_3 + \dots + X_t + X_{t+1}}{t} = F_{t+1} + \frac{X_{t+1}}{t} - \frac{X_1}{t} =$$
$$= F_{t+1} + \frac{1}{t}(X_{t+1} - X_1)$$

untuk  $t \geq n$ , maka

$$F_{t+2} = F_{t+1} + \frac{1}{n}(X_{t+1} - X_{t-(n-1)})$$

atau

$$F_{t+1} = F_t + \frac{1}{n}(X_t - X_{t-n})$$

jika data  $X_{t-n}$  tidak tersedia, dapat diganti dengan  $F_t$ .

Sehingga persamaan di atas menjadi:

$$F_{t+1} = F_t + \frac{1}{n}(X_t - F_t) = \frac{1}{n}X_t + \left(1 - \frac{1}{n}\right)F_t \text{ atau}$$

$$F_{t+1} = \alpha X_t + (1 - \alpha)F_t$$

### b. Pemulusan Eksponensial Ganda (Metode Brown)

- Untuk data time series yang mengikuti suatu trend linier

$$S'_t = \alpha X_t + (1 - \alpha)S'_{t-1}$$

$$S''_t = \alpha S'_t + (1 - \alpha)S''_{t-1}$$

$$a_t = S'_t + (S'_t - S''_t) = 2S'_t - S''_t$$

$$b_t = \frac{\alpha}{1 - \alpha}(S'_t - S''_t)$$

$$F_{t+m} = a_t + b_t m$$

### c. Pemulusan Eksponensial Ganda:

#### Metode 2 Parameter dari Holt

- Untuk data trend
- Memperhalus trend dan slope-nya secara langsung dengan menggunakan konstanta penghalusan yang berbeda

$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + b_{t-1})$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}$$

$$F_{t+m} = S_t + b_t m$$

**d. Pemulusan Eksponensial Tripel:  
Metode 3 Parameter dari Winter**

- Untuk data time series yang mengandung unsur trend dan musiman

Pemulusan keseluruhan:

$$S_t = \alpha \frac{X_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1})$$

Pemulusan trend:

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}$$

Pemulusan musiman:

$$I_t = \beta \frac{X_t}{S_t} + (1 - \beta)I_{t-L}$$

$$F_{t+m} = (S_t + b_t m)I_{t-L+m}$$

**Ukuran Kesalahan Dalam Peramalan:**

• **Kegunaan:**

- a. Mengukur keandalan teknik peramalan
- b. Membandingkan ketelitian 2 teknik peramalan
- c. Menentukan teknik peramalan yang optimal

**a. Mean Absolute Deviation (MAD):**

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - F_t|$$

**b. Mean Square Error (MSE):**

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - F_t)^2$$

**c. Mean Percentage Error (MPE):**

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - F_t)}{Y_t}$$

**d. Mean Absolute Percentage Error (MAPE):**

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - F_t|}{Y_t}$$

Contoh:

Diketahui Data Bulanan Perjalanan Penumpang Udara selama 5 tahun:

|         | Jan | Peb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|         | Nov | Dec |     |     |     |     |     |     |     |     |
| Tahun-1 | 171 | 180 | 193 | 181 | 183 | 218 | 230 | 242 | 209 | 191 |
| Tahun-2 | 172 | 194 |     |     |     |     |     |     |     |     |
| Tahun-3 | 196 | 196 | 236 | 235 | 229 | 243 | 264 | 272 | 237 | 211 |
| Tahun-4 | 180 | 201 |     |     |     |     |     |     |     |     |
| Tahun-5 | 204 | 188 | 235 | 227 | 234 | 264 | 302 | 293 | 259 | 229 |
|         | 203 | 229 |     |     |     |     |     |     |     |     |
|         | 242 | 233 | 267 | 269 | 270 | 315 | 364 | 347 | 312 | 274 |
|         | 237 | 278 |     |     |     |     |     |     |     |     |

|  |     |     |     |     |     |     |     |     |     |     |
|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|  | 284 | 277 | 317 | 313 | 318 | 374 | 413 | 405 | 355 | 306 |
|  | 271 | 306 |     |     |     |     |     |     |     |     |

**Hasil Peramalan Dengan Minitab, Untuk Berbagai Metode:**

**Single Exponential Smoothing**

Data                    PPU  
 Length                60.0000  
 NMissing             0

Smoothing Constant  
 Alpha: 1.43489

Accuracy Measures

MAPE:    8.417  
 MAD:    21.505  
 MSD:    649.237

| Row | Period | Forecast | Lower   | Upper   |
|-----|--------|----------|---------|---------|
| 1   | 61     | 325.357  | 272.671 | 378.043 |
| 2   | 62     | 325.357  | 272.671 | 378.043 |
| 3   | 63     | 325.357  | 272.671 | 378.043 |
| 4   | 64     | 325.357  | 272.671 | 378.043 |

**Double Exponential Smoothing**

Data                    PPU

Length            60.0000  
NMissing        0

Smoothing Constants

Alpha (level): 1.42799  
Gamma (trend): 0.03680

Accuracy Measures

MAPE:     8.762  
MAD:     22.233  
MSD:    682.648

| Row | Period | Forecast | Lower   | Upper   |
|-----|--------|----------|---------|---------|
| 1   | 61     | 327.002  | 272.531 | 381.472 |
| 2   | 62     | 328.689  | 217.439 | 439.939 |
| 3   | 63     | 330.376  | 161.545 | 499.208 |
| 4   | 64     | 332.063  | 105.458 | 558.669 |

**Winters' multiplicative model**

Data            PPU  
Length        60.0000  
NMissing     0

Smoothing Constants

Alpha (level): 0.2

Gamma (trend): 0.2

Delta (seasonal): 0.2

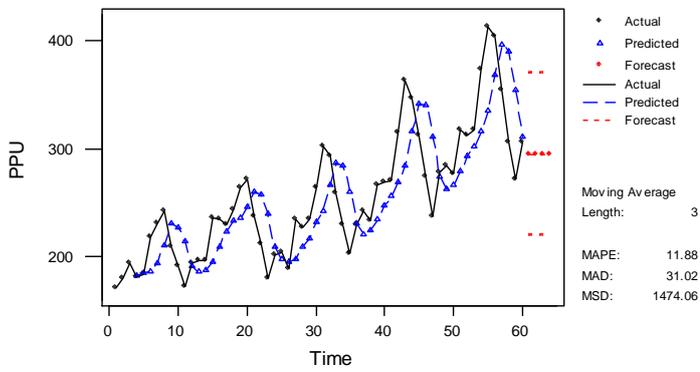
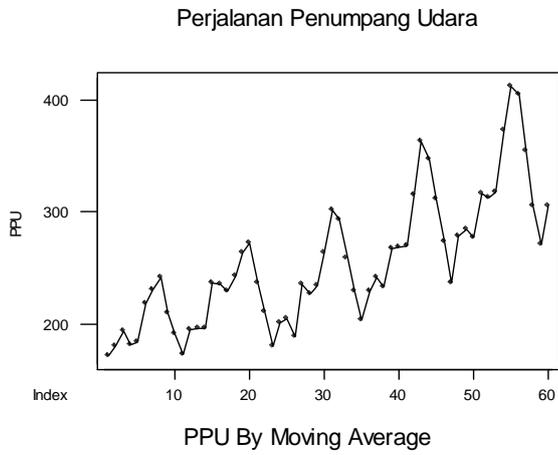
Accuracy Measures

MAPE: 13.43

MAD: 34.68

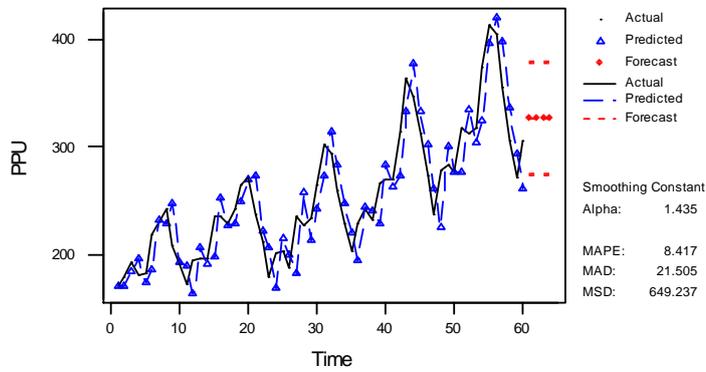
MSD: 1768.12

| Row | Period | Forecast | Lower   | Upper   |
|-----|--------|----------|---------|---------|
| 1   | 61     | 342.502  | 257.546 | 427.459 |
| 2   | 62     | 322.180  | 235.454 | 408.906 |
| 3   | 63     | 340.549  | 251.916 | 429.182 |
| 4   | 64     | 342.133  | 251.466 | 432.799 |

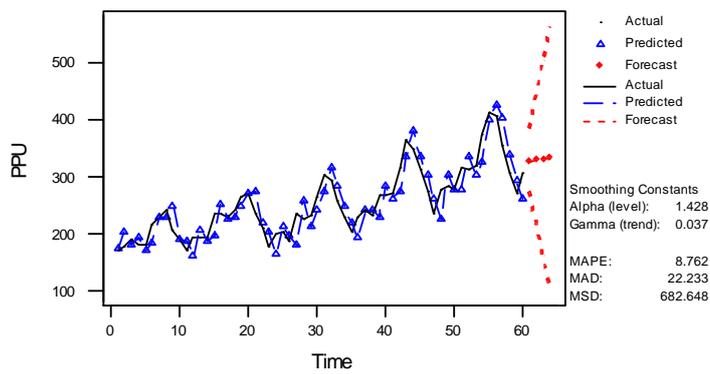


**Plot:**

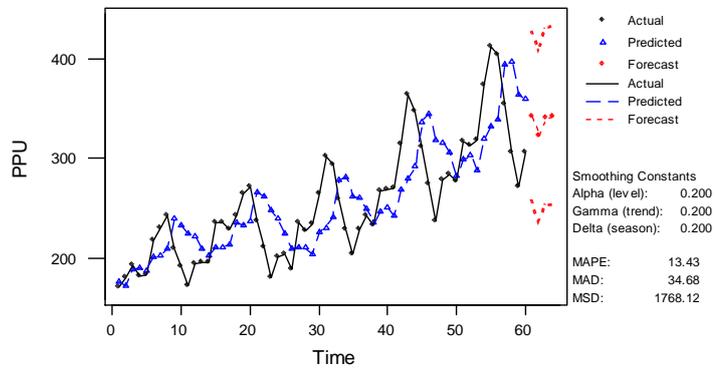
PPU By Single Exp.Smoothing



PPU By Double Exp.Smoothing



PPU By Winter's Method



### C. TUGAS PENDAHULUAN

Carilah contoh kasus yang menggunakan peramalan dengan time series. ( 5 contoh)

### D. PERCOBAAN

| Period | Demand |
|--------|--------|
| 1      | 37     |
| 2      | 40     |
| 3      | 41     |
| 4      | 37     |
| 5      | 45     |

|    |    |
|----|----|
| 6  | 50 |
| 7  | 43 |
| 8  | 47 |
| 9  | 56 |
| 10 | 52 |
| 11 | 55 |
| 12 | 54 |

Gunakan Metode Exponential Smoothing untuk meramal demand untuk tahun ke 13 dan hitung MAD nya

### **E. LAPORAN RESMI**

Kumpulkan hasil percobaan di atas , tambahkan dalam laporan resmi, analisa dan aplikasinya.

# PRAKTIKUM 15/16

---

## STUDI KASUS DI INDUSTRI\_1

---

### A. TUJUAN PEMBELAJARAN

Mengaplikasikan proses pengambilan data, hasil kuisioner di bidang industri / perusahaan.

### B. DASAR TEORI

#### 1. Perumusan masalah

Dalam perumusan masalah ada beberapa yang perlu kita mengerti adalah sebagai berikut :

- permasalahan utama sehingga perlu dilakukan penelitian,
- tujuan dilaksanakannya penelitian,
- cara yang digunakan sehingga datanya bisa diperoleh,
- kemampuan untuk melakukan penelitian dilihat dari biaya, tenaga, waktu dan latar belakang teori,
- ijin penelitian,
- jumlah informasi yang sudah kita peroleh ,serta
- masih perlunya dilakukan studi pendahuluan .

#### 2. Landasan teori

Dari landasan teori maka dapat ditentukan variabel dan sumber data. Variabel yang akan diteliti adalah gejala yang nilainya bervariasi. Gejala yang nilainya selalu tetap tidak dapat digunakan sebagai variabel penelitian.

### **3. Desain perumusan hipotesisnya**

Perumusan Hipotesis dalam penelitian sangat diperlukan, sebelum kita mencapai suatu kesimpulan kita akan melakukan dugaan akan variabel apa saja yang terkait terhadap suatu masalah.

### **4. Skala pengukurannya**

Dalam penelitian sangat diperlukan skala pengukuran.

### **5. Jumlah sample yang diperlukan**

Hal-hal yang diperlukan dalam pengambilan sample sebagai berikut:

- a. Seberapa besar keragaman populasi
- b. Berapa besar tingkat keyakinan yang kita perlukan
- c. Berapa toleransi tingkat kesalahan dapat diterima
- d. Apa tujuan penelitian yang akan dilakukan
- e. Keterbatasan yang dimiliki oleh peneliti

### **6. Teknik pengambilan sampel**

Berkaitan dengan pengumpulan data

1. Data apa saja yang harus dikumpulkan ?
2. Bagaimana instrumen untuk mengumpulkan data ?
3. Siapa yang akan mengumpulkan data ?
4. Berapa biaya untuk mengumpulkan data ?
5. Berapa tenaga yang diperlukan untuk mengumpulkan data ?
6. Bagaimana prosedur yang harus dipenuhi untuk mengumpulkan data ?

Syarat-syarat data yang baik adalah Data harus Akurat, Data harus relevan dan Data harus *up to date*.

**Teknik Pengumpulan Data :**

- 1) Angket (Kuesioner),
- 2) Pengamatan (Observation).
- 3) Wawancara (Interview),

**D. PERCOBAAN**

Lakukan penelitian tentang “ Memprediksi talenta mahasiswa PENS di dunia kerja“

Yang harus dilakukan :

1. cari variable factor- faktornya ( dengan kusioner, wawancara atau angket)
2. Lakukan korelasi antara variable tersebut dengan talenta mahasiswa.
3. Mencari Data pendukung
4. Selidikilah apakah ada daa outlier?
5. Lakukan peramalan.

**E. LAPORAN RESMI**

Kumpulkan hasil percobaan di atas , tambahkan dalam laporan resmi analisa kasus tersebut.

# DAFTAR PUSTAKA

1. Abdusy Syarif ST., MT, Tata Cara Dasar Melakukan Survei di Internet, Pusat Pengembangan Bahan Ajar - UMB.
2. Darman, SE.,MM, Teknik wawancara, PUSAT PENGEMBANGAN BAHAN AJAR-UMB
3. Widya Setiabudi , modul 1,2006
4. Nur Iriawan, PhD. Pemodelan TIME SERIES STATISTIKA UNTUK TRANSPORTASI ,2001
5. DeRETBERKALA (*TIME SERIES*), 2007
6. Materi Statistik Dasar
7. Pertemuan 9 – Data Time Series (Deret Waktu)
8. *Kasmir, jakfar*. Studi Kelayakan Bisnis. 2003. Jakarta : Prenada Media