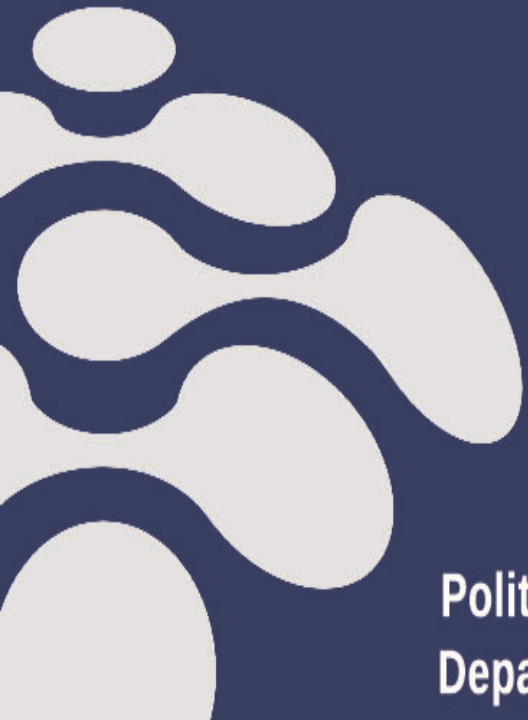


Manajemen Pemrosesan Data (MPD)

Data Identification and Data Cleansing

Oleh Politeknik Elektronika Negeri Surabaya

2018



Politeknik Elektronika Negeri Surabaya
Departemen Teknik Informatika dan Komputer

Topik Materi Hari Ini

- OPEN DATA – **Open Data ID vs Open Data TW**
- DATA IDENTIFICATION
- DATA CLEANSING
- ANALYZING : **Time Series Data**
- EXTRACTING : **Trend, Season, and Variance**

Open Data ID VS Open Data TW

Open Data – First sight



Open Data – First sight



Open Data – First sight

臺北市府資料開放平台

搜尋 資料目錄 開發應用 使用規範 最新消息 常見問答 會員專區 相關連結 關於網站 ⓘ 系統問題回報

首頁 > 資料目錄 > 主題

Taiwan ranks No. 1 in global open data survey

主題分類瀏覽

- 農業
- 民政
- 文化
- 經濟
- 教育
- 環保
- 族群
- 財稅

The Importance of OPEN DATA

Usefulness of Data

The Importance of OPEN DATA

Quantity

The Importance of OPEN DATA

Quality

The Importance of OPEN DATA

Periodic

The Importance of OPEN DATA

Consistency

The Importance of OPEN DATA

The variety of provided data

The Importance of OPEN DATA

How deep you can dig on it ?

Sample data of “Traffic Accident”

51686	105,12,27,14,05,2,07信義區,信義區基隆路1段156號,0,1,1,B01,1,53,3,6,50,14,09
51687	105,12,27,14,05,2,07信義區,信義區基隆路1段156號,0,1,2,C03,2,34,2,6,50,14,09
51688	105,12,29,09,55,2,07信義區,信義區基隆路1段與松隆路口,0,1,1,B03,1,43,3,7,50,3,01
51689	105,12,29,09,55,2,07信義區,信義區基隆路1段與松隆路口,0,1,2,C03,2,38,2,7,50,3,01
51690	105,11,27,17,18,2,07信義區,信義區松隆路與永吉路30巷102弄口,0,1,1,A02,1,61,3,8,50,14,08
51691	105,11,27,17,18,2,07信義區,信義區松隆路與永吉路30巷102弄口,0,1,2,H02,2,40,2,8,50,14,08
51692	105,12,27,08,12,2,07信義區,信義區信義路5段,0,1,1,A04,1,35,3,8,50,14,09
51693	105,12,27,08,12,2,07信義區,信義區信義路5段,0,1,2,H02,2,64,2,8,50,14,09

105 years.csv

1	發生年,發生月,發生日,發生時,發生分,處理別,區序,肇事地點,死亡人數,受傷人數,當事人序,車種,性別
2	105.04.07.08.15.2.01大同區,大同區民族西路與延平北路4段口,0,1,2,C03,1,26,2,8,50,5

Sample data of “Traffic Accident”

```

51686 105,12,27,14,05,2,07信義區,信義區基隆路1段156號,0,1,1,B01,1,53,3,6,50,14,09
51687 105,12,27,14,05,2,07信義區,信義區基隆路1段156號,0,1,2,C03,2,34,2,6,50,14,09
51688 105,12,29,09,55,2,07信義區,信義區基隆路1段與松隆路口,0,1,1,B03,1,43,3,7,50,3,01
51689 105,12,29,09,55,2,07信義區,信義區基隆路1段與松隆路口,0,1,2,C03,2,38,2,7,50,3,01
51690 105,11,27,17,18,2,07信義區,信義區松隆路與永吉路30巷102弄口,0,1,1,A02,1,61,3,8,50,14,08
51691 105,11,27,17,18,2,07信義區,信義區松隆路與永吉路30巷102弄口,0,1,2,H02,2,40,2,8,50,14,08
51692 105,12,27,08,12,2,07信義區,信義區信義路5段,0,1,1,A04,1,35,3,8,50,14,09
51693 105,12,27,08,12,2,07信義區,信義區信義路5段,0,1,2,H02,2,64,2,8,50,14,09

```

105 years.csv	
1	發生年,發生月,發生日,發生時,發生分,處理別,區序,肇事地點,死亡人數,受傷人數,當事人序,車種,性別
2	105,04,07,08,15,2,01大同區,大同區民族西路與延平北路4段口,0,1,2,C03,1,26,2,8,50,5

Year of occurrence, Month, Date of occurrence, When it occurs, Occurrence, Handle, Zonal order, Anecdote location, death toll, Number of injured, Party order, Car model, gender, age, Degree of injury, the weather, Speed limit, Road type, Accident location

Sample data of “Traffic Accident”

```

51686 105,12,27,14,05,2,07信義區,信義區基隆路1段156號,0,1,1,B01,1,53,3,6,50,14,09
51687 105,12,27,14,05,2,07信義區,信義區基隆路1段156號,0,1,2,C03,2,34,2,6,50,14,09
51688 105,12,29,09,55,2,07信義區,信義區基隆路1段與松隆路口,0,1,1,B03,1,43,3,7,50,3,01
51689 105,12,29,09,55,2,07信義區,信義區基隆路1段與松隆路口,0,1,2,C03,2,38,2,7,50,3,01
51690 105,11,27,17,18,2,07信義區,信義區松隆路與永吉路30巷102弄口,0,1,1,A02,1,61,3,8,50,14,08
51691 105,11,27,17,18,2,07信義區,信義區松隆路與永吉路30巷102弄口,0,1,2,H02,2,40,2,8,50,14,08
51692 105,12,27,08,12,2,07信義區,信義區信義路5段,0,1,1,A04,1,35,3,8,50,14,09
51693 105,12,27,08,12,2,07信義區,信義區信義路5段,0,1,2,H02,2,64,2,8,50,14,09

```

105 years.csv

```

1 發生年,發生月,發生日,發生時,發生分,處理別,區序,肇事地點,死亡人數,受傷人數,當事人序,車種,性別
2 105,04,07,08,15,2,01大同區,大同區民族西路與延平北路4段口,0,1,2,C03,1,26,2,8,50,5

```

Year of occurrence, Month, Date of occurrence, When it occurs, Occurrence, Handle, Zonal order, Anecdote location, death toll, Number of injured, Party order, Car model, gender, age, Degree of injury, the weather, Speed limit, Road type, Accident location

IDE APA YANG DAPAT ANDA IMAJINASIKAN DENGAN CONTOH DATA SEPERTI INI ?

INFORMASI APA SAJA YANG DAPAT DI EKSTRAK ?

How make the data ready to use ?

1	Year of occurrence,	Month,	Date of occurrence,	When it occurs,	Occurrence
2	104,1,1,0,18,2,08	Shilin District,	Dongshan Road;	Lane 25; No. 097;	k
3	104,1,1,0,18,2,08	Shilin District,	Dongshan Road;	Lane 25; No. 097;	k
4	104,1,1,0,18,2,08	Shilin District,	Dongshan Road;	Lane 25; No. 097;	k
5	104,1,1,0,18,2,08	Shilin District,	Dongshan Road;	Lane 25; No. 097;	k
6	104,1,1,1,37,2,10	Wenshan District,	4; Muzha Road;	Xiu Ming Road,0,2,	
7	104,1,1,1,37,2,10	Wenshan District,	4; Muzha Road;	Xiu Ming Road,0,2,	
8	104,1,1,1,37,2,10	Wenshan District,	4; Muzha Road;	Xiu Ming Road,0,2,	
9	104,1,1,1,37,2,10	Wenshan District,	4; Muzha Road;	Xiu Ming Road,0,2,	
0	104,1,1,2,30,2,06	Songshan District,	Tayou Road No. 6 in the water ga		
1	104,1,1,2,30,2,06	Songshan District,	Tayou Road No. 6 in the water ga		
2	104,1,1,2,30,2,06	Songshan District,	Tayou Road No. 6 in the water ga		

Make it “convenient” and “readable”

How good is your data ?

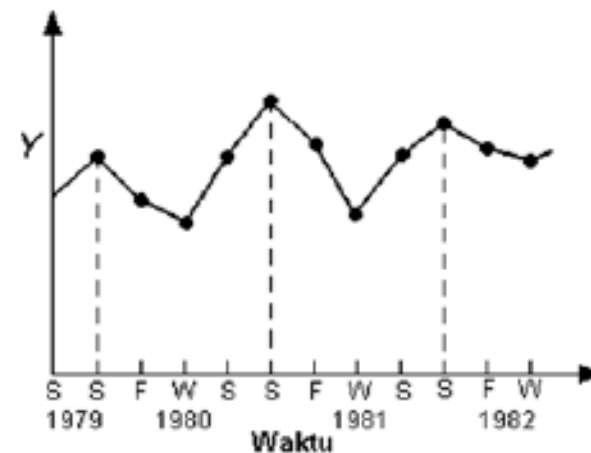
Year of occurrence																
	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
oc	When it o	Occurrence	Handle	Zonal orde	Anecdote	death toll	Number o	Party orde	Car model	gender	age	Degree of	the weath	Speed ??I	Road type	Accident loca
1	0	18	2	08 Shilin C	Dongshan	0	3	1	C03	1	19	2	8	30	15	9
1	0	18	2	08 Shilin C	Dongshan	0	3	2	C03	1	18	2	8	30	15	9
1	0	18	2	08 Shilin C	Dongshan	0	3	3	C03	1	19	3	8	30	15	9
1	0	18	2	08 Shilin C	Dongshan	0	3	4	H02	2	20	2	8	30	15	9
1	1	37	2	10 Wensha 4; Muzha f		0	2	1	B03	1	51	3	8	50	4	1
1	1	37	2	10 Wensha 4; Muzha f		0	2	2	C03	1	19	2	8	50	4	1
1	1	37	2	10 Wensha 4; Muzha f		0	2	3	H02	2	18	2	8	50	4	1
1	1	37	2	10 Wensha 4; Muzha f		0	2	4	H02	1	13	4	8	50	4	1
1	2	30	2	06 Songsh Tayou Roa		0	2	1	C03	1	19	2	8	0	3	22
1	2	30	2	06 Songsh Tayou Roa		0	2	2		3			8	0	3	22
1	2	30	2	06 Songsh Tayou Roa		0	2	3	H02	2	18	2	8	0	3	22
1	2	35	2	03 Zhongs 4; Muzha f		0	1	1	G06	1		4	7	50	14	9
1	2	35	2	03 Zhongs 4; Muzha f		0	1	2	C03	1	18	3	7	50	14	9
1	2	35	2	03 Zhongs 4; Muzha f		0	1	3	H02	2	18	2	7	50	14	9
1	2	45	2	08 Shilin C	4; Muzha f	0	1	1	B02	1	22	3	8	50	4	1
1	2	45	2	08 Shilin C	4; Muzha f	0	1	2	C03	1	22	2	8	50	4	1
1	2	50	2	08 Shilin C	4; Muzha f	0	1	1	C03	1	18	2	7	50	4	1
1	2	50	2	08 Shilin C	4; Muzha f	0	1	2	B01	1	50	3	7	50	4	1
1	3	8	2	03 Zhongs 4; Muzha f		0	1	1	C03	1	19	2	7	40	3	2
1	3	8	2	03 Zhongs 4; Muzha f		0	1	2		3			7	40	3	2
1	3	25	2	09 Beitou 4; Muzha f		0	1	1	B03	2	28	3	8	60	14	9
1	3	25	2	09 Beitou 4; Muzha f		0	1	2	C03	1	18	2	8	60	14	9



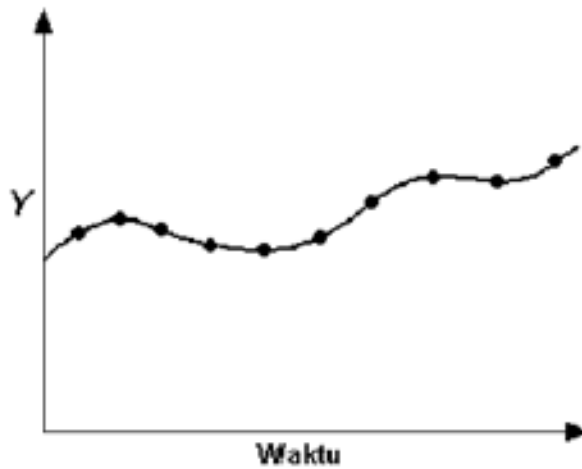
Pola Data



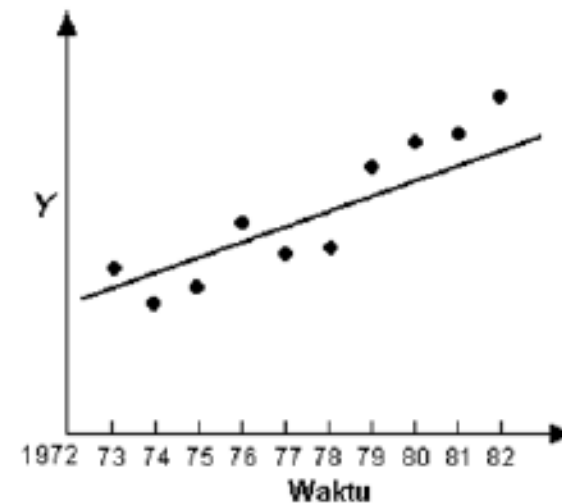
Gambar 6.1. Pola pola Horizontal



Gambar 6.2. Pola Data Musiman



Gambar 6.3. Pola pola Siklis



Gambar 6.4. Pola Data Trend

Trend Data

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

Varian Data - Variasi Musim

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.



Tugas MPD Online 11

1. Download raw data **Traffic Accident** di:
http://bit.ly/rawdata_traffic_accd
2. Urut sesuai NRP. Pemilihan DataSet: Tahun 102, 103, ..., 106.
3. Tiap individu memilih 1 dataset
4. Lakukan proses Identification dan Cleaning pada dataset anda!
5. Analisa informasi apakah yang memungkinkan untuk di ekstrak ?
6. Visualisasikan menjadi 4 Macam Informasi (chart)!
7. Buat Laporan Resmi.
8. Deadline **@Sebelum TM** upload di:
<https://s.id/mpd-ak17-m11>



Tugas MPD Online 12

1. Download File **Praktikum 6** di: http://bit.ly/mpd_p6
2. Gunakan **dataset** anda yang digunakan pada Hasil tugas **MPD Online 11**, setelah dataset diidentifikasi dan di-cleaning
3. Lakukan data ekstraksi informasi data menjadi data **per-triwulan** untuk **NRP GENAP** dan **per-dua bulan** untuk **NRP GANJIL**.
4. Lakukan sesuai praktikum 6:
 1. Pahami dan analisa deret waktu dari dataset tersebut
 2. Cari trend, Season, dan Varian pada dataset tsb. (Setiap kelompok tidak harus memiliki output luaran yang sama)
5. Buat Laporan Resmi.
6. Deadline **@Sebelum TM** upload di:
<https://s.id/mpd-ak17-m12>



bridge to the future

<http://www.eepis-its.edu>

