

REPORT

Exploratory Data Analysis : Preprocessing Data

(Live Editor Tasks)



by :

KAIRA MILANI FITRIA

2221210065

MAGISTER TEKNIK INFORMATIKA

IIB DARMAJAYA

BANDAR LAMPUNG

2023

Table of contents:

1. Clean Missing Data _____
2. Clean Outlier Data _____
3. Find Change Points _____
4. Find Local Extrema _____
5. Smooth Data _____
6. Remove Trends _____
7. Join Tables _____

Source:

1 – 6 :

https://www.mathworks.com/help/matlab/data_analysis/cleandatawithliveeditortasks.html

7 :

<https://www.mathworks.com/help/matlab/ref/jointables.html>

1. Clean Missing Data

Tahap pertama yang dilakukan adalah membuat dan memplot vektor data yang masih acak, yang berisi empat nilai NaN dan lima outlier.

```
x = 1:100;  
data = cos(2*pi*0.05*x+2*pi*rand) + 0.5*randn(1,100);  
data(20:20:80) = NaN;  
data(10:20:90) = [-50 40 30 -45 35];
```

Proses cleaning data dimulai dengan plotting data, dengan ketentuan

$x = 1:100$, artinya $x = 1$ hingga 100 dengan fungsi data yaitu

$\cos(2\pi \cdot 0.05 \cdot x + 2\pi \cdot \text{rand}) + 0.5 \cdot \text{randn}(1, 100)$

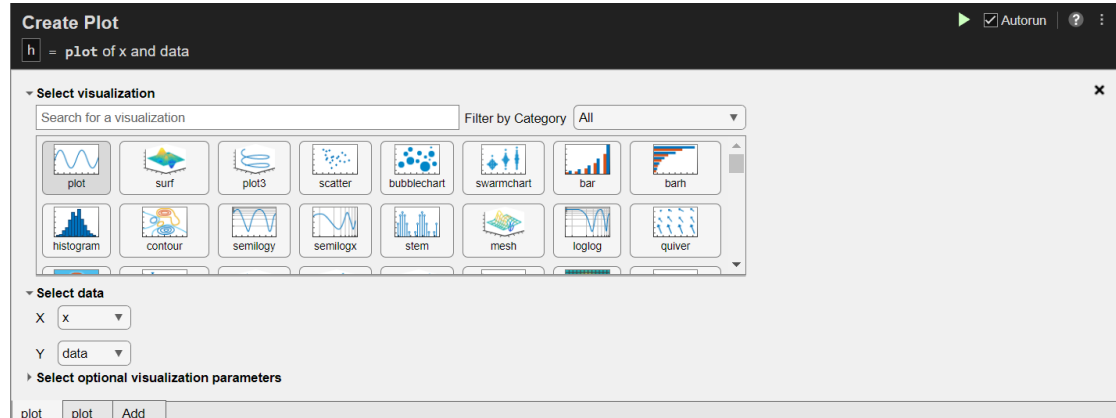
Kemudian ditambahkan ketentuan pada data 20 sampai 80 setiap kelipatan 20 (20 40 60 80) data kosong atau NaN

$\text{data}(20:20:80) = \text{NaN};$

dan pada data 10 sampai 90 setiap kelipatan 20 (10 30 50 70 90) nilainya (-50 40 30 -45 35).

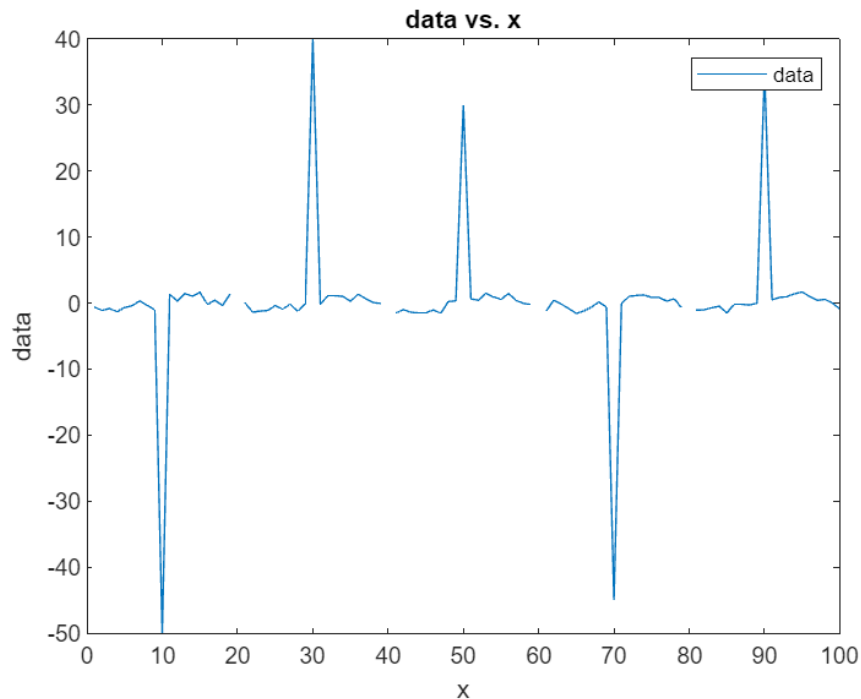
$\text{data}(10:20:90) = [-50 \ 40 \ 30 \ -45 \ 35];$

Setting untuk plotting dapat dilihat pada Gambar 1.



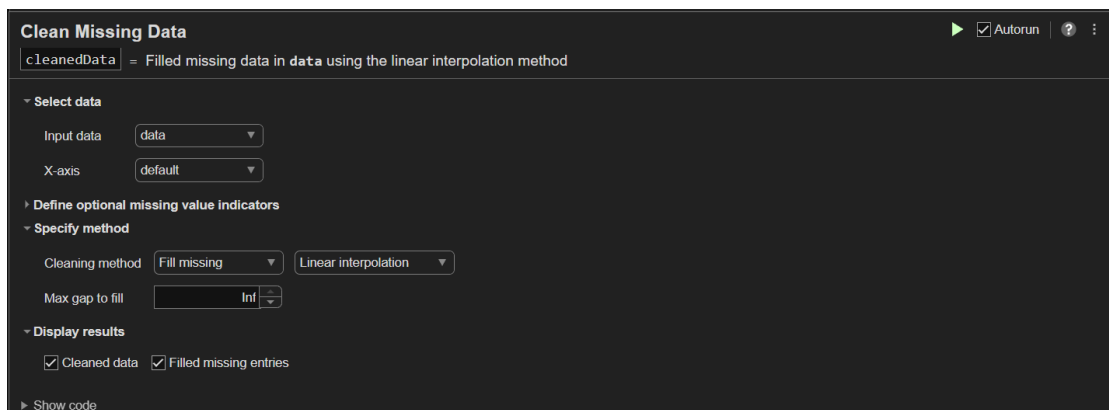
Gambar 1. Plotting data

Hasil dari plotting data dapat dilihat pada Gambar 2.



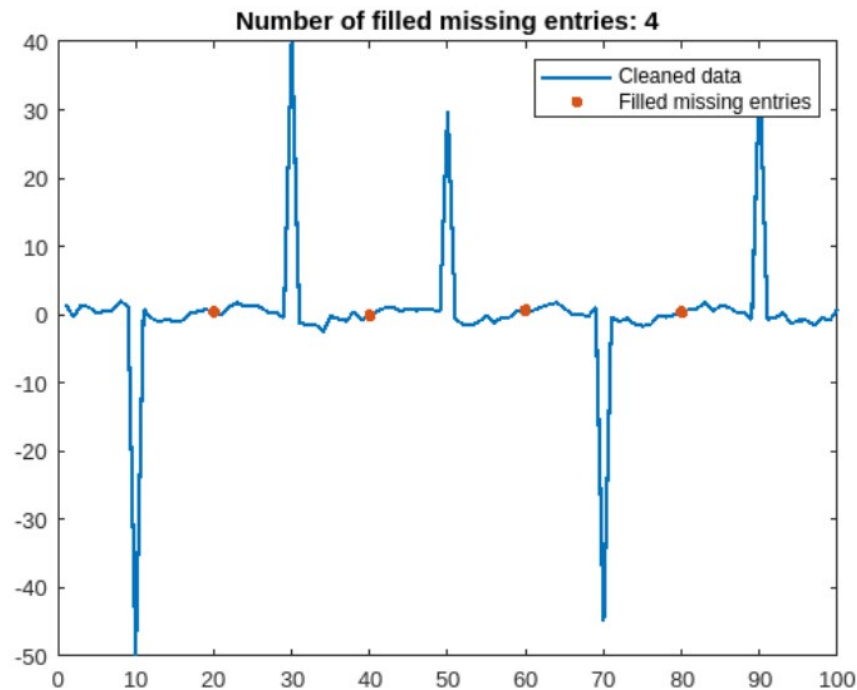
Gambar 2. Hasil plotting nilai x dan data

Tahap kedua, yaitu masuk di task clean missing data, hal ini dilakukan untuk mengganti nilai NaN dalam data dan memvisualisasikan hasilnya, dengan membuka “Clean Missing Data”. Kemudian mengetikkan kata kunci “missing” dalam blok kode, lalu klik “Clean Missing Data” ketika muncul di menu. Pilih data input dan metode pembersihan untuk memplot data yang terisi secara otomatis.



Gambar 3. Clean missing data

Pada mode default terisi cleaning methodnya dengan Fill missing dan Linear interpolation.

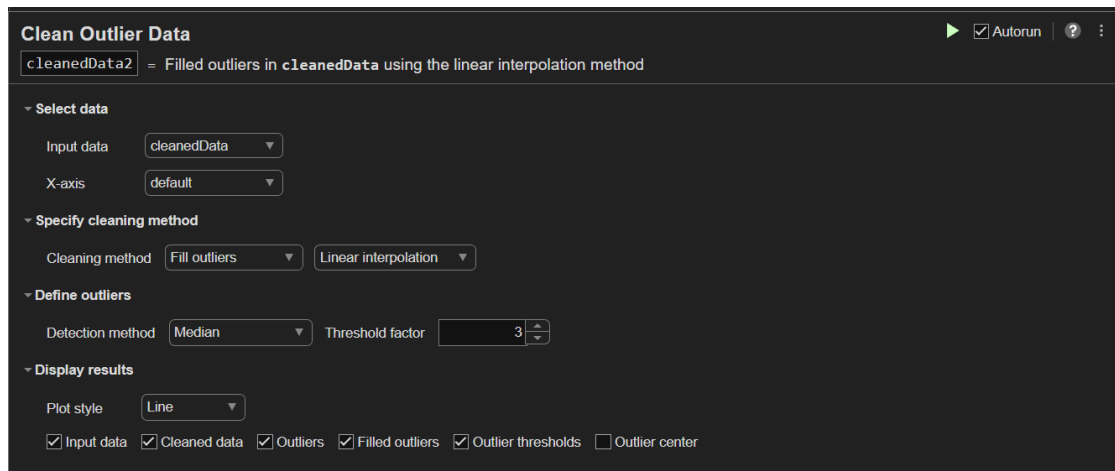


Gambar 4. Hasil clean missing data

Pada Gambar 4 terlihat bahwa nilai yang kosong sebelumnya telah terisi dan ditandai dengan titik berwarna oranye. Perbedaan antara Gambar 2 dan Gambar 4 adalah garis pada Gambar 4 lebih tebal dibandingkan dengan Gambar 2. Terdapat 4 titik nilai kekosongan yang berhasil diisi dengan menggunakan metode Fill Missing Data. Nilai kosong ini terisi karena merujuk kepada aturan $\text{data}(20:20:80) = \text{NaN}$.

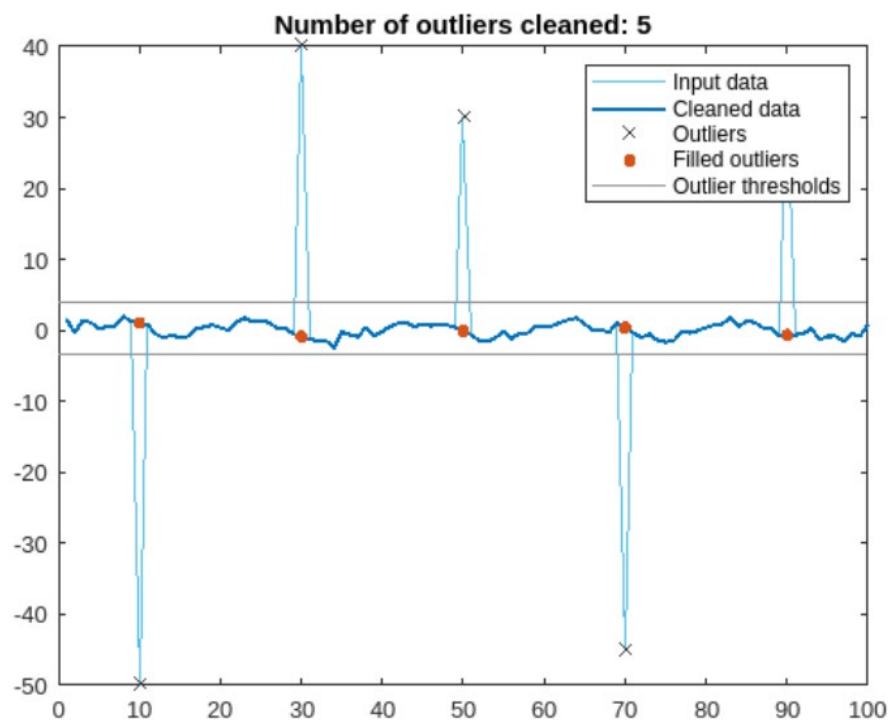
2. Clean Outlier Data

Penghapusan data outlier dari data yang dibersihkan (cleaned data) yang telah dibersihkan pada tugas sebelumnya dengan menggunakan "Clean Outlier Data". Penulisan keyword outlier di blok kode baru dan klik "Clean Outlier Data" untuk membuka task tersebut. Dengan memilih "cleanedData" sebagai data input. Tujuan utama dari tugas ini adalah untuk menghilangkan data yang dianggap ekstrim atau terlalu jauh dari rata-rata nilai yang ada kemudian menggantinya dengan nilai yang berada di ranah rata-rata yang ada.



Gambar 5. Clean outlier data

Pada threshold 3 nilai outliers yang didapatkan ada 5, 5 disini merupakan titik yang lokasinya berada diluar batas threshold ditandai dengan x. Selanjutnya nilai x disesuaikan nilainya agar masuk ke dalam batas threshold dan bergabung dengan kurva data x. Penyesuaian data ini menggunakan metode linear interpolation.

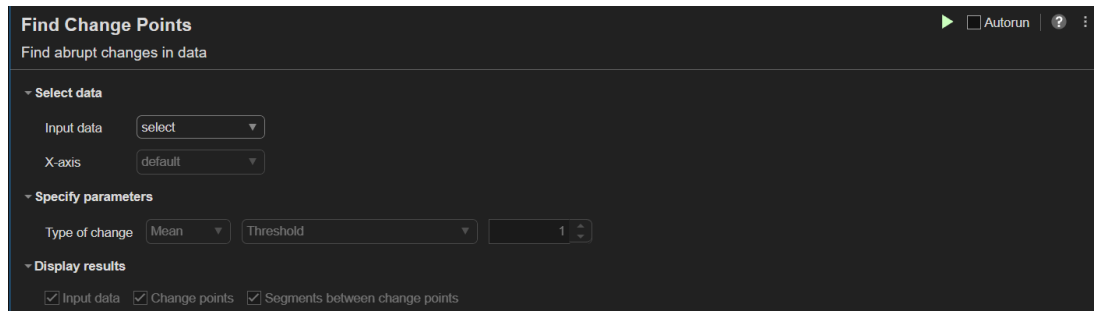


Gambar 6. Hasil clean outlier data

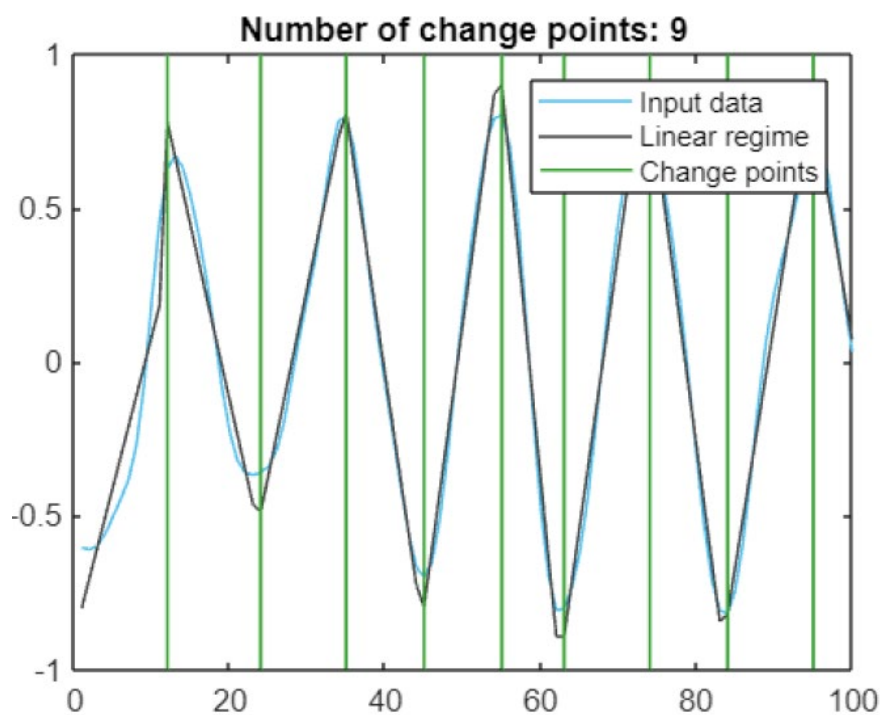
3. Find Change Points

Task find change points menentukan banyaknya perubahan pada kurva data x. Task Find Change Points memungkinkan secara interaktif menemukan perubahan mendadak pada mean, varians, atau kemiringan

dan intersep data. Konfigurasinya adalah input data yang digunakan adalah smoothed data atau data yang telah melewati smoothing. Pada menu type of change digunakan metode linear dengan mode threshold dengan nilai 1. Hasilnya didapatkan 9 perubahan sesuai dengan grafik kurva data x.



Gambar 7. Task find change points



Gambar 8. Hasil find change points

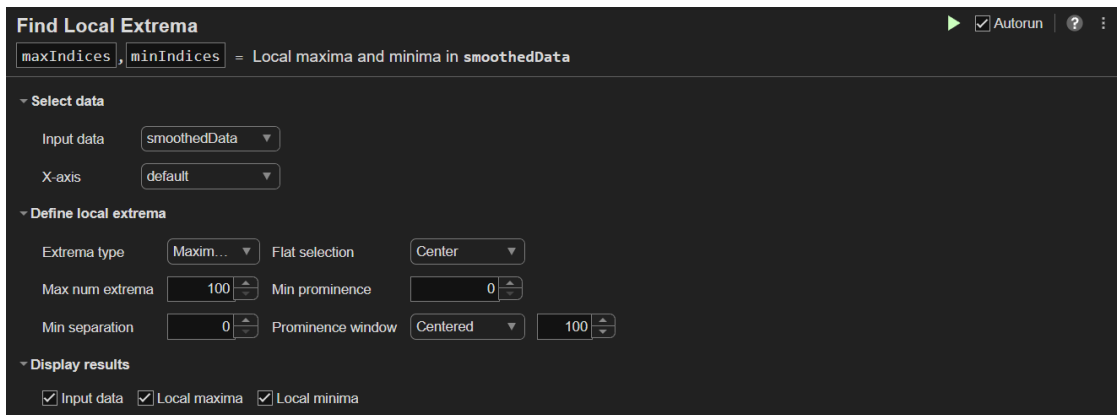
Fitur ini berfungsi untuk mendeteksi berapa banyak perubahan nilai data yang tersedia, baik itu terjadi peningkatan maupun penurunan data. Fitur ini dapat mendeteksi adanya penyimpangan ketika dalam suatu rentang rata-rata data ada yang nilainya signifikan.

4. Find Local Extrema

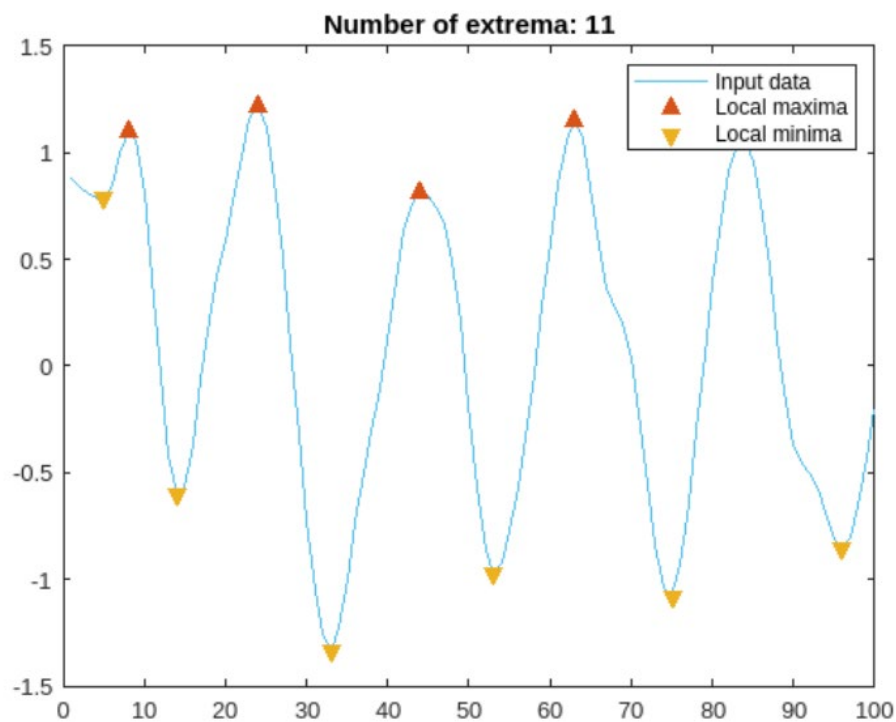
Dengan mengetikkan kata kunci extrema dan klik Find Local Extrema, kemudian gunakan smoothedData sebagai data input dan ubah jenis extrema untuk menemukan maxima lokal dan minima lokal dari data yang

telah dibersihkan dan dihaluskan. Dapat dilakukan penyesuaian parameter extrema lokal untuk menemukan lebih banyak atau lebih sedikit maxima dan minima.

Kurva yang telah dihaluskan selanjutnya dilakukan pemetaan nilai-nilai puncak dan juga lembahnya. Hal ini digunakan untuk melihat atau menandai posisi nilai tertinggi dan terendah pada kurva.



Gambar 9. Find local extrema



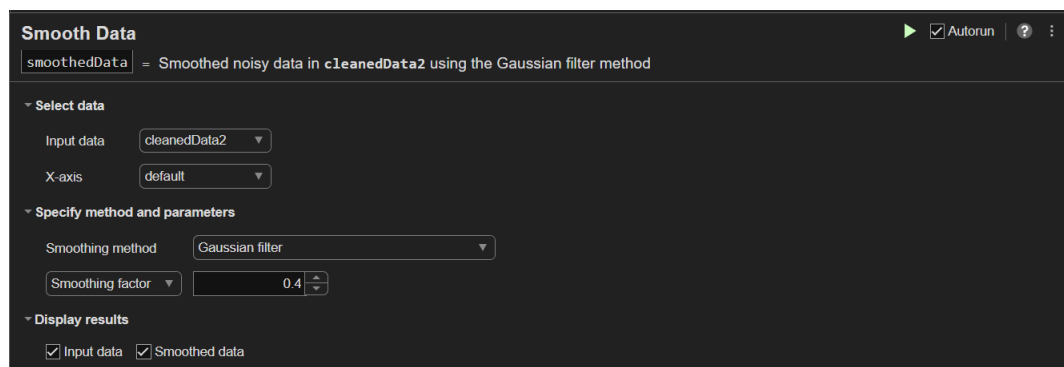
Gambar 10. Hasil find local extrema

Nilai maxima and minima agar nilai yang tampil nanti berupa nilai tertinggi dan terendah pada kurva. Kemudian nilai num extreme dipilih 100 dan minimumnya 0. Kemudian untuk menu min separation dibuat 0 karena kami ingin setiap nilai tertinggi dan terendah pada kurva agar ditandai dan dijumlahnya diketahui pada hasil. Number of extrema ini didapatkan dari

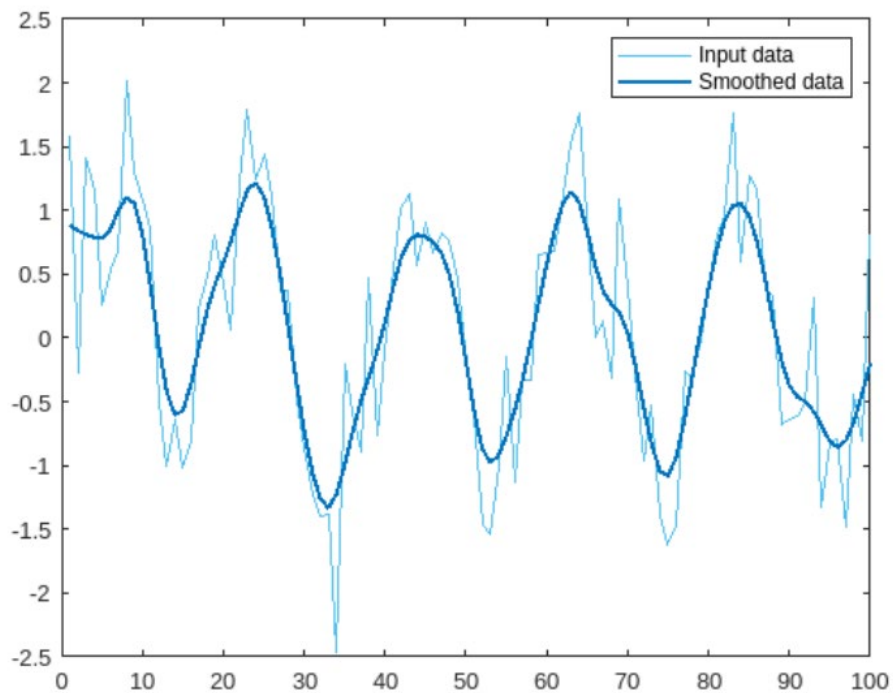
jumlah nilai maxima yang bukan 0 dengan nilai minima yang nilainya bukan 0.

5. Smooth Data

Selanjutnya, perbaiki data yang telah dibersihkan dari tugas sebelumnya dengan menggunakan Smooth Data. Ketik kata kunci smooth dan klik task tersebut ketika muncul. Pilih cleanedData2, output dari tugas sebelumnya, sebagai data input. Pilih metode smoothing, dan sesuaikan faktor smoothing untuk lebih banyak atau lebih sedikit faktor smoothing. Pada mode default digunakan metode Gaussian dengan tingkat kehalusan atau smooth 0.4. Beberapa metode penghalusan kurva ini digunakan jika masih terdapat nilai Outliers pada kurva seperti metode Moving median, Robust Lowess, dan juga Robust Loess. Pada menu Smoothing faktor akan berfungsi jika nilai-nilai pada kurva semakin tidak beraturan maka smoothing factor ini harus semakin ditingkatkan nilainya. Fungsinya untuk mencari nilai trend yang paling representatif pada setiap titik untuk menghasilkan kurva yang semakin jelas polanya



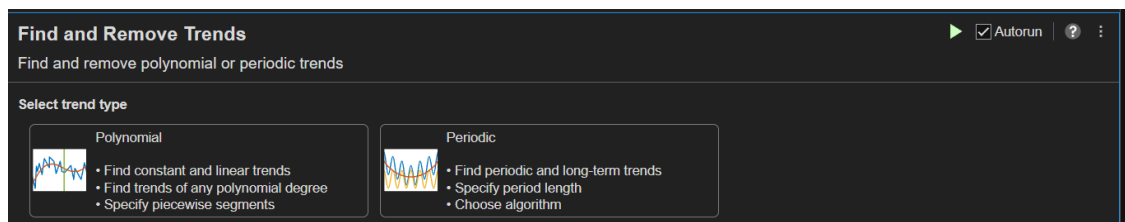
Gambar 11. Smooth Data



Gambar 12. Hasil smooting data

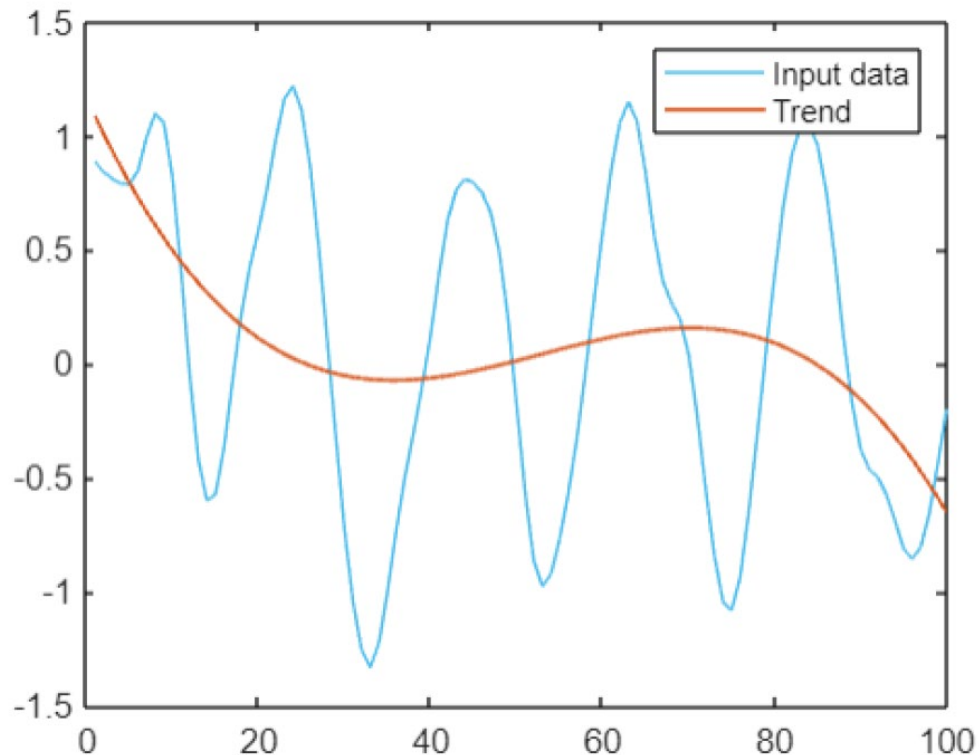
6. Remove Trends

Dari task find local extrema didapatkan nilai maxima dan minima, selanjutnya akan dilakukan penambahan trend pada kurva data yang telah melewati smoothing. Pada input data digunakan smoothedData dengan metode polynomial type menggunakan cubic.



Gambar 13. Task remove trends

Trend yang didapatkan berdasarkan pola grafik yang ada pada data x. Dengan metode cubic hasil trend yang didapatkan sangat baik jika dibandingkan dengan metode yang lain, sebab polanya bisa mengikuti pola grafik yang ada pada data x.



Gambar 14. Hasil remove trends

7. Join Tables

Tugas Join Tables memungkinkan untuk menggabungkan dua tabel secara interaktif dengan melakukan penggabungan atau dengan menggabungkan tabel secara horizontal atau vertikal. Task ini juga dapat untuk melakukan penggabungan, termasuk penggabungan dalam dan luar, pada tabel dengan menentukan variabel kunci-variabel tabel yang nilainya cocok, atau nilai kunci, menentukan baris mana yang digabungkan dalam tabel keluaran. Task Join tables juga dapat menggabungkan tabel secara horizontal atau vertikal dengan jumlah baris atau variabel yang sama.

Contoh : Join Tables Using Live Editor Task

Pertama, muat tabel order, yang berisi nomor order, nama pelanggan, dan tanggal order untuk sejumlah pengiriman.

```
load orders
orders
```

orders = 3x3 table

	OrderID	Custom...	OrderD...
1	5120	"Sanchez"	23-Apr-2019
2	1037	"Li"	18-Apr-2019
3	8937	"Johnson"	16-Apr-2019

Kemudian muat tabel items, yang berisi produk yang dipesan pelanggan, beserta jumlah, harga, dan status pengiriman untuk item tersebut. Setiap baris dari tabel ini memiliki order ID, sama seperti pesanan. Karena

pelanggan dapat memesan beberapa barang, beberapa baris barang dapat merujuk ke satu pesanan dari pesanan.

```
load items
items
```

items = 5x5 table

	OrderID	Product	Quantity	Price	Status
1	6005	"Dozen Ro...	1	39.9900	Shipped
2	1037	"Petunia Ba...	1	23.9900	Delivered
3	5120	"Tulips"	12	0.9900	Pending
4	1037	"Gardenias"	1	17.9900	Shipped
5	1037	"Gerber Dai...	6	1.9900	Delivered

Buka task Join Tables (Gabungkan Tabel). Untuk membuka task, ketik kata kunci join di blok kode dan pilih Join Tables saat muncul di menu. Task ini untuk melakukan penggabungan pesanan dan item. Saat tugas terbuka:

1. Pilih orders dan items sebagai tabel kiri dan kanan.
2. Pilih OrderID sebagai variabel penggabungan untuk kedua tabel.
3. Klik tombol Inner join.
4. Untuk melihat kode yang dihasilkan oleh tugas ini, perluas tampilan tugas dengan mengeklik Show code di bagian bawah area parameter tugas.

Join Tables ▶ ✓ Autorun ? :

`joinedData` = Combine orders and items using inner join

▼ Select data

Left table: Right table:

Merging variable: Merging variable: +

▼ Specify join

☒ Outer join
 ☐ Left outer join
 ☐ Right outer join
 ☒ Inner join
 ☐ Join

▼ Display results

☐ Input tables ☒ Output table

joinedData = 4x7 table

	OrderID	Custom...	OrderD...	Product	Quantity	Price	Status
1	1037	"Li"	18-Apr-2019	"Petunia Ba...	1	23.9900	Delivered
2	1037	"Li"	18-Apr-2019	"Gardenias"	1	17.9900	Shipped
3	1037	"Li"	18-Apr-2019	"Gerber Dai...	6	1.9900	Delivered
4	5120	"Sanchez"	23-Apr-2019	"Tulips"	12	0.9900	Pending

Gambar 15. Task join tables (inner join)

Selanjutnya, gunakan task untuk melakukan outerjoin left. Outerjoin dapat menyertakan nilai kunci yang hanya muncul dalam satu tabel input.

1. Klik tombol Left outerjoin.
2. Pilih kotak centang Combine merging variables. Secara default, outer join menyalin variabel kunci dari tabel kiri dan kanan ke dalam variabel terpisah di tabel keluaran. Gabungkan variabel kunci sehingga satu variabel kunci ada di output.

3. Untuk melihat kode yang dihasilkan oleh tugas ini, perluas tampilan tugas dengan mengeklik Show code di bagian bawah area parameter tugas.

Join Tables ▶ ☒ Autorun ? :

`joinedData2` = Combine `orders` and `items` using left outer join

Select data

Left table: `orders` Right table: `items`

Merging variable: `OrderID` Merging variable: `OrderID` +

Specify join

☒ Outer join ☐ Left outer join ☐ Right outer join ☐ Inner join ☐ Join

☒ Combine merging variables

Display results

☐ Input tables ☒ Output table

`joinedData2` = 5x7 table

	OrderID	Custom...	OrderD...	Product	Quantity	Price	Status
1	1037	"Li"	18-Apr-2019	"Petunia Ba ...	1	23.9900	Delivered
2	1037	"Li"	18-Apr-2019	"Gardenias"	1	17.9900	Shipped
3	1037	"Li"	18-Apr-2019	"Gerber Dai ...	6	1.9900	Delivered
4	5120	"Sanchez"	23-Apr-2019	"Tulips"	12	0.9900	Pending
5	8937	"Johnson"	16-Apr-2019	<missing>	NaN	NaN	<undefined>

Gambar 16. Task join tables (left outer join)