

LAPORAN
RENCANA TUGAS MANDIRI (RTM) Ke-3
MATA KULIAH MACHINE LEARNING I
“MENERAPKAN EXPONENTIAL SMOOTHING PADA DATASET
HARIAN COVID-19 TAHUN 2020-2022”



DISUSUN OLEH:

Muhammad Aryasatya Nugroho (22083010085)

DOSEN PENGAMPU:

Tresna Maulana Fahrudin S.ST., M.T. (NIP. 199305012022031007)

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU, Asean. Eng. (NIP. 198012052005011000)

PROGRAM STUDI SAINS DATA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR

2024

Exponential Smoothing

Exponential smoothing adalah salah satu metode yang digunakan dalam analisis *time series* untuk meramalkan data dengan memberikan bobot yang lebih besar pada data terbaru. Metode ini sangat berguna dalam meramalkan data yang memiliki tren, musiman, atau fluktuasi acak.

Istilah "eksponensial" mengacu pada fakta bahwa dalam proses *smoothing*, bobot yang diberikan pada pengamatan data berkurang secara eksponensial seiring berjalannya waktu. Dengan kata lain, data terbaru memiliki pengaruh yang lebih besar daripada data yang lebih lama.

Teknik ini memiliki parameter yang disebut *alpha*, yang mengontrol seberapa besar dampak pengamatan terbaru terhadap ramalan. Nilai *alpha* yang lebih tinggi memberikan lebih banyak bobot pada data terbaru, sementara nilai *alpha* yang lebih rendah memberikan bobot yang lebih merata pada data historis.

Exponential smoothing cocok untuk data *time series* yang memiliki tren linier atau nonlinier, serta data dengan pola musiman. Namun, perlu diingat bahwa *exponential smoothing* memiliki asumsi bahwa data yang diamati terdiri dari sinyal dan *noise*, dan bahwa sinyal berubah secara eksponensial.

Dalam praktiknya, *exponential smoothing* sering digunakan dalam berbagai bidang, termasuk ekonomi, keuangan, peramalan penjualan, dan pengelolaan rantai pasokan. Metode ini relatif sederhana dan mudah diterapkan, namun dapat memberikan hasil yang cukup akurat tergantung pada kecocokannya dengan karakteristik data yang diamati.

Cara Menghitung Exponential Smoothing

Langkah-langkah cara menghitung *exponential smoothing* sebagai berikut:

- a) Menghitung Koefisien α
- b) Menghitung nilai peramalan periode pertama
- c) Menghitung nilai peramalan di keseluruhan periode
- d) Menghitung akurasi peramalan dengan peta kontrol *Tracking Signal*
- e) Membuat grafik peta control *Tracking Signal*

Cara menghitung *Exponential Smoothing* dengan memberikan bobot secara eksponensial atau bertingkat pada data-data terbaru, sehingga data terbaru tersebut akan mendapatkan bobot yang lebih besar. Metode *Exponential smoothing* digunakan untuk menghitung data dengan

menggunakan permintaan data aktual beberapa bulan yang telah berlalu dimana data cenderung tidak stabil dan berubah-ubah setiap waktu.

Rumus Exponential Smoothing

Rumus *exponential smoothing* adalah sebagai berikut:

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

Dimana:

F_t : Nilai ramalan untuk periode ke-t

A_{t-1} : Nilai aktual periode ke-t

F_{t-1} : Nilai ramalan untuk periode waktu lalu; t-1

α : Konstanta pemulusan (*smoothing constant*)

Catatan

- Apabila data permintaan sangat bergejolak atau tidak stabil dari waktu ke waktu maka dengan memilih nilai α yang mendekati 1. seperti 0,9 atau 0,8, dst
- Apabila data permintaan relatif stabil atau tidak bergejolak dari waktu ke waktu maka dengan memilih nilai α yang mendekati 0, seperti 0,1 atau 0,2 dst
- Untuk angka pertama dalam rumus, diperoleh dari rata-rata data permintaan aktual dibagi 12 bulan

Di dalam menentukan nilai α dengan cara *trial and error* sesuai catatan diatas, dapat juga dilakukan dengan menghitung rumus $\alpha = (2/n+1)$

Contoh Studi Kasus Exponential Smoothing

Diberikan dataset covid harian pada tahun 2020 hingga tahun 2022. Dataset dapat dilihat seperti dibawah

No	Periode	Aktual
0	1 2020-02-03 00:00:00	2
1	2 2020-03-03 00:00:00	0
2	3 2020-04-03 00:00:00	0
3	4 2020-05-03 00:00:00	0
4	5 2020-06-03 00:00:00	2
...
924	925 2022-12-09 00:00:00	1848
925	926 9/13/2022	2896
926	927 9/14/2022	2799
927	928 9/15/2022	2651
928	929 9/16/2022	2358

929 rows × 3 columns

Dataset diatas mencatat jumlah kasus COVID-19 selama rentang waktu dari awal 2020 hingga akhir 2022. Dengan jumlah kasus yang tercatat pada setiap metode, dataset ini dapat digunakan untuk menganalisis tren penyebaran COVID-19, memperkirakan dampak kebijakan penanggulangan yang diambil oleh pemerintah, atau mempelajari pola musiman atau fluktuasi dalam penyebaran virus.

Dataset tersebut memiliki 929 baris (dari baris 0 hingga baris 928) dan tiga kolom yang terdiri dari kolom 'No' merupakan identifikasi unik untuk setiap periode, kolom 'Periode' untuk tanggal atau waktu masing-masing periode, dan 'Aktual' merupakan jumlah aktual kasus COVID-19 yang tercatat pada setiap periode.

1. Penyelesaian Menggunakan Excel

EXPONENTIAL SMOOTHING										
[0]	[1]	[2]	[3]	[4] = [2] - [3]	[5] = Kumulatif [4]	[6] = ABS([4])	Kumulatif ABS Error [7] = Kumulatif [6]	[8] = [7] / [0]	[9] = [5] / [8]	[10] = [6] / [2] * 100
No	Periode	Aktual	Forecast	Error	RSFE	Absolute Error	Kumulatif Absolut Error	MAD	Tracking Signal	MAPE
1	03/02/2020	2	6894,557589	-6892,557589	-6892,557589	6892,557589	6892,557589	6892,557589	-1	344627,8794
2	03/03/2020	0	1380,511518	-1380,511518	-8273,069107	1380,511518	8273,069107	4136,534553	-2	#DIV/0!
3	03/04/2020	0	276,1023036	-276,1023036	-8549,17141	276,1023036	8549,17141	2849,723803	-3	#DIV/0!
4	03/05/2020	0	55,22046071	-55,22046071	-8604,391871	55,22046071	8604,391871	2151,097968	-4	#DIV/0!
5	03/06/2020	2	11,04409214	-9,044092142	-8613,435963	9,044092142	8613,435963	1722,687193	-5	452,2046071
6	03/07/2020	0	3,808818428	-3,808818428	-8617,244781	3,808818428	8617,244781	1436,207464	-6	#DIV/0!
7	03/08/2020	2	0,761763686	1,238236314	-8616,006545	1,238236314	8618,483018	1231,21186	-6,397988589	61,91181572
8	03/09/2020	13	1,752352737	11,24764726	-8604,758898	11,24764726	8629,730665	1078,716333	-7,976850478	86,52036356
9	03/10/2020	8	10,75047055	-2,750470547	-8607,509368	2,750470547	8632,481136	959,1645706	-8,973965086	34,38088184
10	03/11/2020	7	8,550094109	-1,550094109	-8609,053462	1,550094109	8634,03123	863,403123	-9,971077511	22,14420156

919	09/06/2022	3607	2448,40863	1158,59137	-4399,094829	1158,59137	991018,4801	1078,366137	-4,079407426	32,12063683
920	09/07/2022	3513	3375,281726	137,7182741	-4261,376555	137,7182741	991156,1984	1077,343694	-3,955447624	3,920246914
921	09/08/2022	3138	3485,456345	-347,4563452	-4608,8329	347,4563452	991503,6548	1076,5512	-4,28110888	11,07254127
922	09/09/2022	2804	3207,491269	-403,491269	-5012,324169	403,491269	991907,146	1075,8212	-4,659068041	14,38984554
923	09/10/2022	2609	2884,698254	-275,6982538	-5288,022423	275,6982538	992182,8443	1074,954328	-4,919299627	10,56720022
924	09/11/2022	1939	2664,139651	-725,1396508	-6013,162073	725,1396508	992907,9839	1074,57574	-5,595847597	37,39760963
925	09/12/2022	1848	2084,02793	-236,0279302	-6249,190003	236,0279302	993144,0119	1073,669202	-5,820405383	12,77207414
926	9/13/2022	2896	1895,205586	1000,794414	-5248,395589	1000,794414	994144,8063	1073,590504	-4,88863824	34,55781816
927	9/14/2022	2799	2695,841117	103,1588828	-5145,236707	103,1588828	994247,9652	1072,543652	-4,797228251	3,685562086
928	9/15/2022	2651	2778,368223	-127,3682234	-5272,60493	127,3682234	994375,3334	1071,525144	-4,920654416	4,804535022
929	9/16/2022	2358	2676,473645	-318,4736447	-5591,078575	318,4736447	994693,807	1070,714539	-5,22181998	13,5060918

Untuk menghitung nilai *forecasting* pada periode pertama dapat dihitung dengan menjumlahkan semua permintaan aktual, kemudian dibagi dengan jumlah atau panjang datanya (pada *dataset* ini 929) sehingga diperoleh nilai **6894,557589** seperti berikut

	[3]
=SUM(C5:C933)/A933	Forecast
	6894,557589

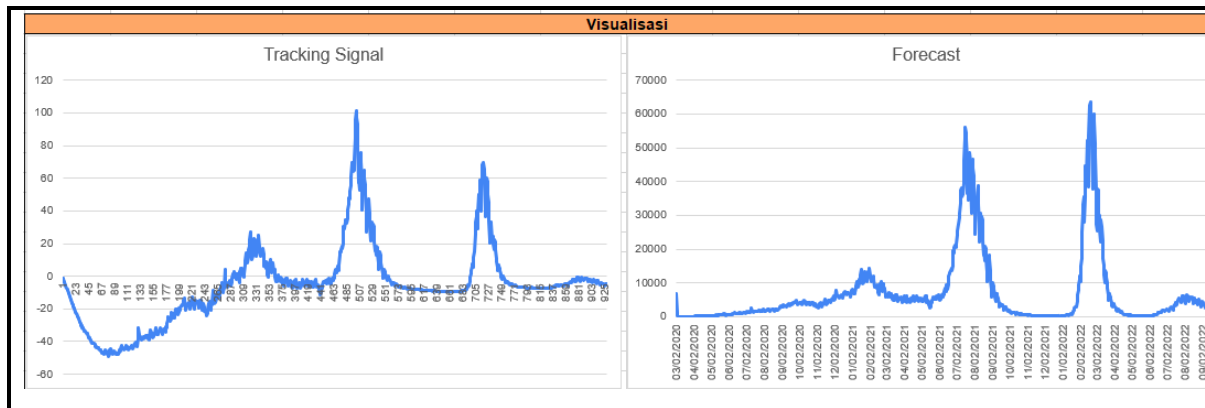
Selanjutnya pada kolom kedua gunakan rumus *exponential smoothing* $F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1})$. Kolom ketiga baris indeks ke-2 mendapatkan hasil **1380,511518** dari perhitungan $F_2 = 6894,557589 + 0,8(2 - 6894,557589)$. Hal ini dilakukan secara kontinu sampai kolom terakhir.

	[3]
	Forecast
	6894,557589
=D5+0,8*(C5-D5)	1380,511518
	276,1023036
	55,22046071
	11,04409214

Menghitung Akurasi Peramalan dengan Peta Kontrol *Tracking Signal*

- Kolom *error* berisi dari pengurangan kolom aktual dengan kolom *forecast*
- Kolom RSFE diperoleh dari kumulatif (penjumlahan) kolom *error*.
- Kolom *absolute error* diperoleh dengan *meng-absolute* kan nilai pada kolom *error*. Yang dimaksud dengan mengabsolutkan adalah nilai negatif berubah menjadi nilai positif dengan rumus excel =abs(kolom error).
- Kolom kumulatif *absolute error* berisi dari kumulatif (penjumlahan) nilai kolom *absolute error*.
- Kolom MAD berasal dari pembagian kolom kumulatif *absolute error* dengan kolom periode.
- Kolom *Tracking signal* diperoleh dengan membagi Kolom RSFE dengan Kolom MAD

Membuat Grafik Peta Kontrol *Tracking Signal* di Excel



Tabel sebelah kiri adalah visualisasi dari hasil perhitungan *tracking signal* dengan interpretasi:

- Secara umum, *tracking signal* mengikuti tren data aktual dengan baik. Hal ini menunjukkan bahwa model *exponential smoothing* mampu memprediksi nilai data dengan cukup akurat.
- Terdapat beberapa periode di mana *tracking signal* sedikit menyimpang dari data aktual. Hal ini dapat disebabkan oleh beberapa faktor, seperti fluktuasi data yang tinggi atau perubahan tren data yang tidak terduga.

Kesimpulan dari visualisasi menunjukkan model *exponential smoothing* dengan α 0,8 dapat digunakan untuk melacak tren data aktual dengan cukup akurat. Model ini dapat membantu dalam membuat prediksi nilai data di masa depan, meskipun terdapat beberapa periode di mana terdapat penyimpangan

Tabel sebelah kanan adalah visualisasi hasil perhitungan *forecasting* dengan interpretasi:

- Secara umum, hasil *forecasting* mengikuti tren data aktual dengan cukup baik. Hal ini menunjukkan bahwa model *forecasting* yang digunakan mampu memprediksi nilai data dengan cukup akurat.
- Terdapat beberapa periode di mana hasil *forecasting* sedikit menyimpang dari data aktual. Hal ini dapat disebabkan oleh beberapa faktor, seperti fluktuasi data yang tinggi atau perubahan tren data yang tidak terduga.
- Pada periode Januari-April 2023, hasil *forecasting* sedikit lebih tinggi daripada data aktual. Hal ini menunjukkan bahwa model *forecasting* mungkin *overestimate* nilai data pada periode tersebut.

- d) Pada periode Mei-Desember 2023, hasil *forecasting* sedikit lebih rendah daripada data aktual. Hal ini menunjukkan bahwa model *forecasting* mungkin *underestimate* nilai data pada periode tersebut.

Kesimpulannya, visualisasi menunjukkan bahwa model perataan eksponensial mampu dengan cukup akurat memprediksi tren data aktual. Meskipun terdapat sedikit penyimpangan antara data aktual dan prediksi model pada beberapa periode, secara umum model ini dapat membantu dalam membuat prediksi nilai data di masa depan. Meskipun demikian, perlu diingat bahwa hasil prediksi dapat bervariasi tergantung pada model yang digunakan dan kualitas data yang tersedia. Oleh karena itu, evaluasi menyeluruh terhadap kinerja model perlu dilakukan sebelum mengandalkan prediksi tersebut untuk pengambilan keputusan.

2. Penyelesaian Menggunakan Python

Load dataset kedalam kode pemrograman python

```
import pandas as pd
data = pd.read_excel("data harian covid-19 2020-2022.xlsx")
data
```

Membuat fungsi untuk menghitung *Exponential Smoothing*

```
def exspsmooth(data,alpha):
    import pandas as pd
    # menghitung forecast
    forecasts = []
    forecast_1 = data['Aktual'].sum() / len(data)
    forecasts.append(forecast_1)
    for i in range(1, len(data)):
        forecast_i = forecasts[i-1] + alpha * (data['Aktual'][i-1] - forecasts[i-1])
        forecasts.append(forecast_i)
    data['Forecast'] = forecasts

    # Menghitung error, RSFE, absolute error, cumulative absolute error
    data['Error'] = data['Aktual'] - data['Forecast']
    data['RSFE'] = data['Error']
    cumulative_rsfe = data['RSFE'].cumsum()
    for i in range(1, len(data)):
        data.at[i, 'RSFE'] = data.at[i, 'Error'] + data.at[i-1, 'RSFE']
    data['Absolute Error'] = abs(data['Error'])
    data['Kumulatif Absolut Error'] = data['Absolute Error'].cumsum()

    # Menghitung MAD
    data['MAD'] = data['Kumulatif Absolut Error'] / data['No']

    # Menghitung tracking signal dan MAPE
    MAD = data['MAD'].mean()
    data['Tracking Signal'] = data['RSFE'] / data['MAD']
    data['MAPE'] = (data['Absolute Error'] / data['Aktual']) * 100

    return data
```

Fungsi 'exspsmooth' pada *script* diatas digunakan untuk menerapkan metode perataan eksponensial pada data *time series*. Berikut penjelasan untuk setiap bagian kodenya

- a) Perhitungan Forecast

Nilai rata-rata dari kolom 'Aktual' dihitung. Selanjutnya, *Forecast* untuk setiap periode dihitung berdasarkan formula perataan eksponensial: $\text{forecast}_i = \text{forecasts}[i-1] + \alpha * (\text{data}['\text{Aktual}'][i-1] - \text{forecasts}[i-1])$, di mana α adalah parameter *smoothing* yang diberikan sebagai input ke fungsi.

b) Perhitungan Error

Setelah *Forecasting* dihitung, *error* dihitung melalui selisih antara nilai aktual dan *forecast* dihitung untuk setiap periode.

c) Perhitungan RSFE (Running Sum of Forecast Errors)

RSFE adalah akumulasi dari semua nilai *error* sebelumnya. RSFE dihitung untuk setiap periode dengan mengakumulasi *error* saat ini dengan RSFE sebelumnya, ini memberikan gambaran akumulatif dari kesalahan prediksi.

d) Perhitungan Kumulatif Absolute Error

Didapatkan dari akumulasi dari semua nilai *error absolute* sebelumnya, ini memberikan gambaran akumulatif dari kesalahan prediksi dalam bentuk nilai *absolute*.

e) Penghitungan MAD (Mean Absolute Deviation)

Didapatkan dari rata-rata seluruh nilai kumulatif *absolute error*

f) Perhitungan *Tracking Signal* dan MAPE (Mean Absolute Error)

Tracking Signal adalah rasio dari RSFE terhadap MAD. Ini memberikan gambaran tentang apakah model overforecasting atau underforecasting. Sedangkan MAPE adalah presentase dari *error* absolut dibandingkan dengan nilai aktual.

Terakhir, fungsi ini mengembalikan DataFrame yang telah diperbarui dengan kolom-kolom baru yang berisi hasil dari perhitungan tersebut.

Memanggil fungsi 'expsmooth(data, alpha)'

```
data = pd.read_excel('data harian covid-19 2020-2022.xlsx')
alpha = 0.8
expsmooth(data, alpha)
```

Digunakan alpha sebesar 0,8 dikarenakan data cenderung sangat bergejolak dan tidak stabil.

Hasil *DataFrame* yang menampilkan perhitungan seluruh kolom

No	Periode	Aktual	Forecast	Error	RSFE	Absolute Error	Kumulatif Absolut Error	MAD	Tracking Signal	MAPE	
0	1	2020-02-03 00:00:00	2	6894.557589	-6892.557589	-6892.557589	6892.557589	6892.557589	6892.557589	-1.000000	3.446279e+05
1	2	2020-03-03 00:00:00	0	1380.511518	-1380.511518	-8273.069107	1380.511518	8273.069107	4136.534553	-2.000000	inf
2	3	2020-04-03 00:00:00	0	276.102304	-276.102304	-8549.171410	276.102304	8549.171410	2849.723803	-3.000000	inf
3	4	2020-05-03 00:00:00	0	55.220461	-55.220461	-8604.391871	55.220461	8604.391871	2151.097968	-4.000000	inf
4	5	2020-06-03 00:00:00	2	11.044092	-9.044092	-8613.435963	9.044092	8613.435963	1722.687193	-5.000000	4.522046e+02
...
924	925	2022-12-09 00:00:00	1848	2084.027930	-236.027930	-6249.190003	236.027930	993144.011865	1073.669202	-5.820405	1.277207e+01
925	926	9/13/2022	2896	1895.205586	1000.794414	-5248.395589	1000.794414	994144.806279	1073.590504	-4.888638	3.455782e+01
926	927	9/14/2022	2799	2695.841117	103.158883	-5145.236707	103.158883	994247.965162	1072.543652	-4.797228	3.685562e+00
927	928	9/15/2022	2651	2778.368223	-127.368223	-5272.604930	127.368223	994375.333386	1071.525144	-4.920654	4.804535e+00
928	929	9/16/2022	2358	2676.473645	-318.473645	-5591.078575	318.473645	994693.807030	1070.714539	-5.221820	1.350609e+01

929 rows × 11 columns

Menghitung Rata-Rata MAPE

```
filter_average = exspsmooth(data,alpha).iloc[1:].replace([np.inf, -np.inf], np.nan).dropna(subset=["MAPE"])
average_mape = filter_average['MAPE'].mean()

print("Rata-rata MAPE:", average_mape)

Rata-rata MAPE: 18.142301491663176
```

Perhitungan Average MAPE diatas mengecualikan baris indeks ke-0 dan hasil perhitungan MAPE yang berupa 'inf'. Diperoleh hasil MAPE sebesar **18.142301491663176%**.

Visualisasi Hasil *Exponential Smoothing*

```
import pandas as pd
import matplotlib.pyplot as plt

data = exspsmooth(data, alpha)

# Visualisasi data aktual dan hasil prediksi
plt.figure(figsize=(12, 6)) # Atur ukuran grafik jika perlu
plt.plot(data['No'], data['Aktual'], label='Aktual', color='red')
plt.plot(data['No'], data['Forecast'], label='Forecast', color='blue')

# Tambahkan grafik tracking signal
plt.axhline(y=1, color='black', linestyle='--', label='Batas Kontrol Atas')
plt.axhline(y=-1, color='black', linestyle='--', label='Batas Kontrol Bawah')
plt.plot(data['No'], data['Tracking Signal'], label='Tracking Signal', color='green')

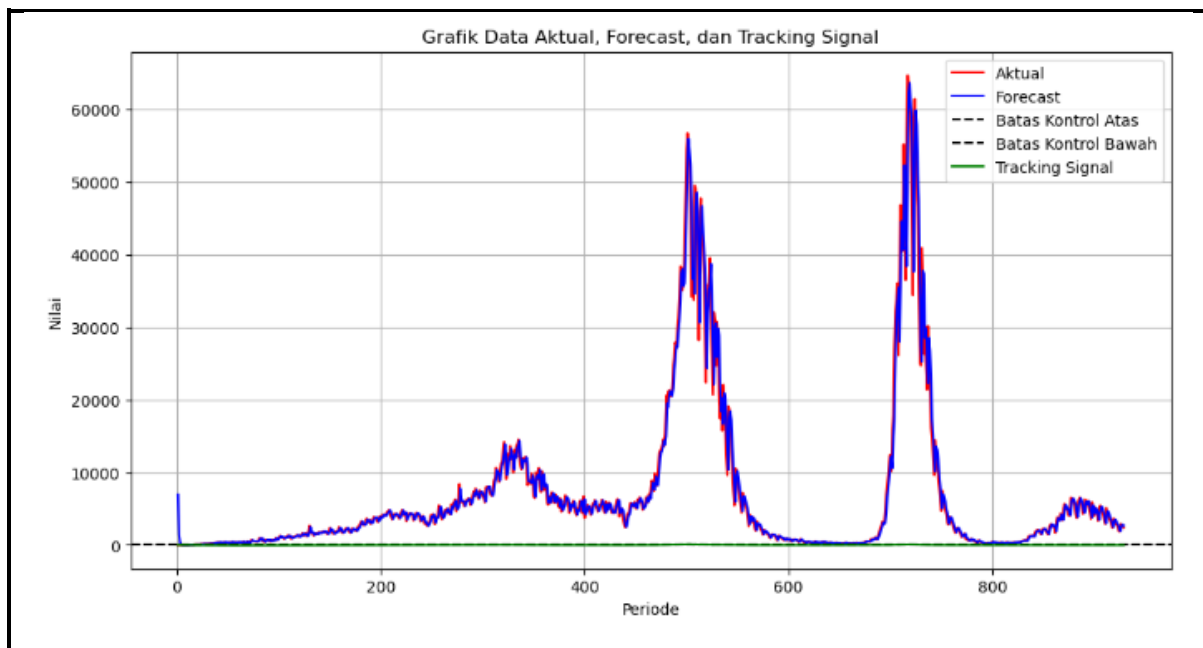
# Atur Label, judul, dan legenda grafik
plt.xlabel('Periode')
plt.ylabel('Nilai')
plt.title('Grafik Data Aktual, Forecast, dan Tracking Signal')
plt.legend()

# Tampilkan grafik
plt.grid(True)
plt.show()
```

Kode visualisasi menggunakan matplotlib untuk menampilkan data aktual, prediksi, dan tracking signal. Pertama, fungsi 'exspsmooth' diterapkan pada data dengan parameter *alpha*.

Plot grafik menampilkan data aktual (merah), prediksi (biru), dan tracking signal (hijau) dengan batas kontrol atas dan bawah pada nilai 1 dan -1. Label sumbu x dan y ditambahkan, serta judul grafik dan legenda. Grid juga ditampilkan untuk mempermudah interpretasi data. Ini memberikan gambaran visual tentang kinerja model dalam memprediksi data COVID-19 dan memberikan peringatan melalui *tracking signal* jika terdapat ketidaksesuaian signifikan antara prediksi dan data aktual.

Hasil Visualisasi



Visualisasi menunjukkan bahwa model *forecasting* yang digunakan cukup akurat dalam memprediksi nilai data. Hal ini dibuktikan dengan tracking signal yang umumnya berada di dalam batas kontrol atas dan bawah. Namun, terdapat beberapa periode di mana model *forecasting* mungkin sedikit *overestimate* atau *underestimate* nilai data. Hal ini perlu dikaji lebih lanjut untuk mengetahui penyebabnya dan meningkatkan akurasi model *forecasting*.

Kesimpulan:

Eksperimen ini menunjukkan bahwa *eksponential smoothing* dapat digunakan untuk meramalkan jumlah kasus COVID-19 dengan cukup akurat. Namun, pemilihan parameter *alpha* yang tepat dan evaluasi kinerja model secara menyeluruh diperlukan untuk mendapatkan prediksi yang lebih akurat.