

**FORECASTING COVID-19 (CONFIRMED AND DEATH) WITH LINIER
REGRESSION, SUPPORT VECTOR MACHINE AND TIME SERIES (AR
AND ARIMA MODEL)**



Disusun Oleh:

1. Yuzarifki Alfan Zuhdhi – 09011381722101 (Ketua + Coding + Analisa)
2. Barzan Trio Putra - 09011381722144 (Laporan + Preprocessing)
3. Muhammad Hafizd - 09011381722128 (Preprocessing + Laporan)
4. M. Taufiq Qurahman - 09011381722092 (Analisa + Preprocessing)
5. Zhafran Farris Ghonim - 09011381722137 (Coding + Analisa)

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA**

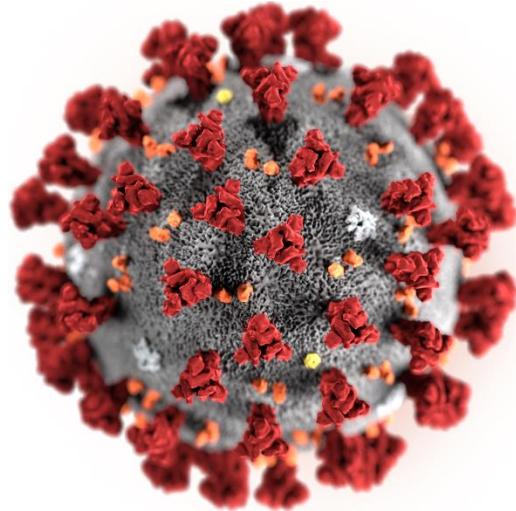
2020

DAFTAR ISI

1. DASAR TEORI	3
2. DATASET	3
3. IMPORTING LIBRARY	3
4. PREPROCESSING	3
5. ANALYSIS DATA	3
6. MACHINE LEARNING.....	3
7. OUTPUT	3

1. DASAR TEORI

Dasar-dasar teori yang menunjang laporan mengenai covid-19, library yang dipakai, tujuan dan sumber coding python.



COVID-19

- COVID-19

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) yang lebih dikenal dengan nama virus Corona adalah jenis baru dari coronavirus yang menular ke manusia. Virus ini bisa menyerang siapa saja, baik bayi, anak-anak, orang dewasa, lansia, ibu hamil, maupun ibu menyusui. Infeksi virus ini disebut *COVID-19* dan pertama kali ditemukan di kota Wuhan, Cina, pada akhir Desember 2019. Virus ini menular dengan cepat dan telah menyebar ke wilayah lain di Cina dan ke beberapa negara, termasuk Indonesia. Hal ini membuat beberapa negara di luar negeri menerapkan kebijakan untuk memberlakukan lockdown dalam rangka mencegah penyebaran virus Corona.

Corona virus adalah kumpulan virus yang bisa menginfeksi sistem pernapasan. Pada banyak kasus, virus ini hanya menyebabkan infeksi pernapasan ringan, seperti flu. Namun, virus ini juga bisa menyebabkan infeksi pernapasan berat, seperti infeksi paru-paru (pneumonia), Middle-East Respiratory Syndrome (MERS), dan Severe Acute Respiratory Syndrome (SARS).

- Gejala Virus Corona (*Covid-19*)

Gejala awal infeksi virus Corona atau COVID-19 bisa berupa gejala flu, seperti demam, pilek, batuk kering, sakit tenggorokan, dan sakit kepala. Setelah itu, gejala bisa memberat. Pasien bisa mengalami demam tinggi, batuk berdahak bahkan berdarah, sesak napas, dan nyeri dada. Gejala-gejala tersebut muncul ketika tubuh bereaksi melawan virus Corona.

Namun, secara umum ada 3 gejala umum yang bisa menandakan seseorang terinfeksi virus Corona, yaitu:

- Demam (suhu tubuh di atas 38 derajat Celsius)
- Batuk
- Sesak Napas

Menurut penelitian, gejala COVID-19 muncul dalam waktu 2 hari sampai 2 minggu setelah terpapar virus Corona.

- Penyebab Virus Corona (*Covid-19*)

Infeksi virus Corona atau COVID-19 disebabkan oleh coronavirus, yaitu kelompok virus yang menginfeksi sistem pernapasan. Pada sebagian besar kasus, coronavirus hanya menyebabkan infeksi pernapasan ringan sampai sedang, seperti flu. Akan tetapi, virus ini juga bisa menyebabkan infeksi pernapasan berat, seperti pneumonia, Middle-East Respiratory Syndrome (MERS), dan Severe Acute Respiratory Syndrome (SARS). Ada dugaan bahwa virus Corona awalnya ditularkan dari hewan ke manusia. Namun, kemudian diketahui bahwa virus Corona juga menular dari manusia ke manusia.

Seseorang dapat tertular COVID-19 melalui berbagai cara, yaitu :

- Tidak sengaja menghirup percikan ludah (droplet) yang keluar saat penderita COVID-19 batuk atau bersin
- Memegang mulut atau hidung tanpa mencuci tangan terlebih dulu setelah menyentuh benda yang terkena cipratan ludah penderita COVID-19
- Kontak jarak dekat dengan penderita COVID-19, misalnya bersentuhan atau berjabat tangan

Virus Corona dapat menginfeksi siapa saja, tetapi efeknya akan lebih berbahaya atau bahkan fatal bila terjadi pada orang lanjut usia, ibu hamil, orang yang memiliki penyakit tertentu, perokok, atau orang yang daya tahan tubuhnya lemah.

LIBRARY

- Pandas

Pandas adalah sebuah librari berlisensi BSD dan open source yang menyediakan struktur data dan analisis data yang mudah digunakan dan berkinerja tinggi untuk bahasa pemrograman Python.

- Matplotlib

adalah librari plotting 2D Python yang menghasilkan gambar publikasi bermutu di dalam berbagai format hardcopy dan lingkungan interaktif sepanjang platform. matplotlib dapat digunakan di dalam script Python, shell Python dan ipython (ala MATLAB®* or Mathematica®), server aplikasi web, dan enam GUI toolkit. matplotlib mencoba untuk membuat hal mudah menjadi lebih mudah dan hal sulit menjadi mungkin. Kamu dapat membuat plot, histogram, power spectra, grafik batang, grafik error, scatterplot, dll, hanya dengan beberapa baris code.

- Numpy

Numpy memiliki kegunaan untuk operasi vektor dan matriks. Fiturnya hampir sama dengan MATLAB dalam mengelola array dan array multidimensi. Numpy merupakan salah satu library yang digunakan oleh library lain seperti Scikit-Learn untuk keperluan analisis data.

- Datetime

Teknik manipulasi data tanggal dan waktu menggunakan modul datetime (pada standard library) yang disediakan python. Keberadaan modul ini sangat mempermudah kita saat harus bekerja dengan tanggal dan waktu dalam kode program kita.

- gridsearchCV

Untuk mencari parameter pengklasifikasikan suatu penyebab yang kamu inginkan. Proses dilakukan secara *brute force* dan melaporkan mana parameter yang memiliki akurasi yang paling baik.

- Linier regression

adalah suatu model statistik yang umum dan paling sederhana yang digunakan untuk Machine Learning untuk melakukan prediksi dengan cara supervised learning.

- Support vector machine

Support Vector Machine (SVM) merupakan sebuah teknik yang menggunakan 2 titik (2 vektor), yang selanjutnya 2 titik ini akan membentuk garis pembatas (atau sisi pembatas jika 3 dimensi atau lebih). Garis/sisi pembatas yang dibentuk dari dua buah vektor ini disebut dengan hyperplane.

- Mean squared error

Mean Squared Error (MSE) atau Mean Squared Deviation (MSD) dari estimator mengukur rata-rata kuadrat kesalahan yaitu perbedaan rata-rata kuadrat antara nilai estimasi dan nilai sebenarnya. Ini adalah fungsi risiko, sesuai dengan nilai yang diharapkan dari hilangnya kesalahan kuadrat. Itu selalu non-negatif dan nilai mendekati nol lebih baik. MSE adalah momen kedua dari kesalahan (tentang asal) dan dengan demikian menggabungkan varians dari estimator dan biasnya.

- ARIMA

ARIMA adalah akronim yang merupakan singkatan dari AutoRegressive Integrated Moving Average. Ini adalah kelas model yang menangkap serangkaian struktur temporal standar yang berbeda dalam data deret waktu.

TUJUAN (GOALS)

- Tujuan dari notebook ini adalah untuk mempelajari wabah COVID-19 dengan bantuan beberapa teknik visualisasi dasar. Perbandingan Cina di mana COVID-19 awalnya berasal dari Rest of the World. Melakukan prediksi dan

perkiraan Time Series untuk mempelajari dampak dan penyebaran COVID-19 di hari-hari mendatang.

SUMBER CODE

- (<https://www.kaggle.com/neelkudu28/covid-19-visualizations-predictions-forecasting>) untuk forecasting dan sumber coding utama.
- (<https://www.kaggle.com/khotijahs1/data-visualisation-indonesia-coronavirus>) untuk visualisasi dan analisis data khusus Indonesia.

2. DATASET

Kami menggunakan 2 dataset, Untuk Forecasting kami gunakan dataset **CSCEGISandData** dari github. Lalu untuk visualisasi dan analisis Negara Indonesia kami gunakan dataset **INDONESIA CORONAVIRUS DATASET**.

CSCEGISandData (<https://github.com/CSSEGISandData/COVID-19>)

- Penjelasan dataset covid_19_data.csv
 - Sno - Serial number
 - ObservationDate - Date of the observation in MM/DD/YYYY
 - Province/State - Province or state of the observation (Could be empty when missing)
 - Country/Region - Country of observation
 - Last Update - Time in UTC at which the row is updated for the given province or country. (Not standardised and so please clean before using it)
 - Confirmed - Cumulative number of confirmed cases till that date
 - Deaths - Cumulative number of deaths till that date
 - Recovered - Cumulative number of recovered cases till that date

INDONESIAN CORONAVIRUS DATASET

(<https://www.kaggle.com/ardisragen/indonesia-coronavirus-cases>)

- Penjelasan cases.csv
 - date = Date
 - new_testedNumber of new tested in this day
 - acc_testedTotal number of tested samples
 - new_confirmedNumber of new confirmed cases in this day
 - acc_confirmedtotal/accumulate of confirmed cases
 - acc_negativetotal/accumulate of negative cases
 - being_checkedtotal/accumulate of sample under examination
 - isolatedtotal/accumulate isolated patient
 - new_releasedNumber of new release cases in this day
 - acc_releasedtotal/accumulate of release cases
 - new_deceasedNumber of new decease/death cases in this day
 - acc_deceasedtotal/accumulate of deceased cases
 - positive_ratenumber of confirmed per number of tested
 - negative_ratenumber of negative per number of tested
 - decease_ratenumber of decease per number of confirmed
 - release_ratenumber of release per number of confirmed
 - dailypositive_ratenumber of new positive per number of new tested in this day

- Jabar.csv

- date = Date
- positive_totalTotal Confirmed Positive Cases in Jabar
- positive_dailydaily confirmed positive cases in Jabar
- odp_processNumber of People Under Monitoring in Process
- odp_endNumber People End Monitoring
- odp_totalTotal Number of People under monitoring
- odp_dailyDaily Number Increase of People under monitoring
- pdp_processNumber of Processed Patient Under Surveillance
- pdp_endNumber of Patient end surveillance
- pdp_totalTotal Number of Patient under suveillance
- pdp_dailyDaily number increase of Patient under surveillance
- total_deceasedTotal Deceased cases
- total_releasedTotal Released Cases
- total_positive_kotabekasiTotal positive case in Kota Bekasi
- total_positive_depokTotal positive case in Kota depok
- total_positive_kabbekasiTotal positive case in kab bekasi
- total_positive_kotabandungTotal positive case in kota Bandung
- total_positive_kotabogorTotal Positive case in Kota Bogor
- total_positive_kabbogorTotal Positive case in Kab Bogor
- total_positive_kotacimahiTotal Positive case in Kota Cimahi
- total_positive_kabbandungTotal Positive case in kab bandung
- total_positive_kabkuninganTotal Positive case in Kab kuningan
- total_positive_kabpurwakartaTotal Positive Case in Kab Purwakarta
- total_positive_kabcirebonTotal Positive case in Kab Cirebon
- total_positive_kabsukabumiTotal Positive case in kab Sukabumi
- total_positive_kabsumedangTotal Positive case in Kab Sumedang
- total_positive_kotatasikmalayaTotal Positive Case in Kota Tasikmalaya
- total_positive_kabciamisTotal Positive case in Kab Ciamis
- total_positive_kabcianjurtotal Positive case in Kab Cianjur
- total_positive_kabgarutTotal Positive case in Kab Garut
- total_positive_kabindramayuTotal Positive Case in Kab Indramayu
- total_positive_kabkarawangTotal Positive case in Kab Karawang
- total_positive_kabmajalengkaTotal Positive case in Kab Majalengka
- total_positive_kabpangandaranTotal Positive case in Kab Pangandaran
- total_positive_kabsubangTotal Positive case in Kab Subang

- total_positive_kabtasikmalayaTotal Positive case in Kab Tasikmalaya
 - total_positive_kotabanjarTotal Positive case in Kota Banjar
 - total_positive_kotacirebonTotal positive case in Kota Cirebon
 - total_positive_kotasukabumiTotal Positive case in Kota Sukabumi
- Jakarta.csv
 - date = Date
 - odp_processNumber of People Under Monitoring
 - odp_endNumber of People End Monitoring
 - odp_totalTotal People In Monitoring
 - pdp_processNumber of Patient Under Surveillance
 - pdp_endNumber of Patient End Surveillance
 - pdp_totalTotal Patient In Surveillance
 - jakbarTotal PDP and ODP in West Jakarta
 - jakpusTotal PDP and ODP in Central Jakarta
 - jakselTotal PDP and ODP in South Jakarta
 - jaktimTotal PDP and ODP in East Jakarta
 - jakutTotal PDP and ODP in North Jakarta
 - outsidejakartaTotal PDP and ODP surrounded Jakarta
 - unknownTotal PDP and ODP in Jakarta (Unknown)
 - confirmedTotal Confirmed Positive Cases
 - releasedTotal Released Cases
 - deceasedTotal Death Cases
 - hospitalizedNumber of Confirmed Cases Under Hospitalized
 - self_isolatedNumber of Confirmed Cases Self Isolation
- Patient.csv
 - patient_idthe ID of the patient (n-th confirmed patient)
 - genderthe gender of the patient
 - agethe age of the patient
 - nationalitythe nationality of patient
 - provincethe region/province of the patient
 - current_statethe city of the patient
 - contacted_withcontacted/get infected by patient id number
 - confirmed_datethe date of confirmation
 - released_datethe date of discharge
 - deceased_datethe date of decease
 - hospitalhospitalized location

3. IMPORTING LIBRARY (STEP 1)

- Library yang digunakan untuk forecasting COVID-19.

```
In [1]: import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import datetime as dt
from datetime import timedelta
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import LinearRegression,Ridge,Lasso
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error,r2_score
import statsmodels.api as sm
from statsmodels.tsa.ar_model import AR
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf,plot_pacf
#pd.set_option('display.float_format', lambda x: '%.6f' % x)
```

4. PRE-PROCESSING (STEP 2)

- Menyiapkan dan memproses dataset yang digunakan.

```
In [2]: covid=pd.read_csv("covid_19_data.csv")
covid.head()
```

```
Out[2]:   SNo ObservationDate Province/State Country/Region Last Update Confirmed Deaths Recovered
0      1      01/22/2020        Anhui Mainland China 1/22/2020 17:00      1.0     0.0     0.0
1      2      01/22/2020       Beijing Mainland China 1/22/2020 17:00     14.0     0.0     0.0
2      3      01/22/2020    Chongqing Mainland China 1/22/2020 17:00      6.0     0.0     0.0
3      4      01/22/2020       Fujian Mainland China 1/22/2020 17:00      1.0     0.0     0.0
4      5      01/22/2020       Gansu Mainland China 1/22/2020 17:00      0.0     0.0     0.0
```

- Dataset yang dipakai untuk forecasting.

```
In [3]: covid.tail()
```

```
Out[3]:   SNo ObservationDate Province/State Country/Region Last Update Confirmed Deaths Recovered
13524 13525      04/09/2020      Wyoming          US 2020-04-09 23:09:19     230.0     0.0     0.0
13525 13526      04/09/2020     Xinjiang Mainland China 2020-04-09 23:09:19     76.0     3.0    73.0
13526 13527      04/09/2020       Yukon        Canada 2020-04-09 23:09:19      7.0     0.0     0.0
13527 13528      04/09/2020      Yunnan Mainland China 2020-04-09 23:09:19     184.0     2.0   173.0
13528 13529      04/09/2020     Zhejiang Mainland China 2020-04-09 23:09:19    1267.0     1.0  1235.0
```

```
In [4]: print("Size/Shape of the dataset: ", covid.shape)
print("Checking for null values:\n", covid.isnull().sum())
print("Checking Data-type of each column:\n", covid.dtypes)

Size/Shape of the dataset: (13529, 8)
Checking for null values:
   SNo          0
ObservationDate      0
Province/State    6562
Country/Region      0
Last Update        0
Confirmed          0
Deaths              0
Recovered           0
dtype: int64
Checking Data-type of each column:
   SNo        int64
ObservationDate    object
Province/State     object
Country/Region     object
Last Update        object
Confirmed         float64
Deaths            float64
Recovered          float64
dtype: object
```

- Menghilangkan kolom "SNo" dan "Province/State" karena mengandung terlalu banyak missing values menggunakan syntax **Drop**.

```
In [5]: covid.drop(["SNo"], 1, inplace=True)
```

- Convert "Observation Date" kedalam format Datetime agar bisa di prosess kedalam koding dengan syntax **datetime**.

```
In [6]: covid["ObservationDate"] = pd.to_datetime(covid["ObservationDate"])
```

- Data terbaru setelah preprocessing

	covid						
Out[7]:	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
0	2020-01-22	Anhui	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
1	2020-01-22	Beijing	Mainland China	1/22/2020 17:00	14.0	0.0	0.0
2	2020-01-22	Chongqing	Mainland China	1/22/2020 17:00	6.0	0.0	0.0
3	2020-01-22	Fujian	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
4	2020-01-22	Gansu	Mainland China	1/22/2020 17:00	0.0	0.0	0.0
...
13524	2020-04-09	Wyoming	US	2020-04-09 23:09:19	230.0	0.0	0.0
13525	2020-04-09	Xinjiang	Mainland China	2020-04-09 23:09:19	76.0	3.0	73.0
13526	2020-04-09	Yukon	Canada	2020-04-09 23:09:19	7.0	0.0	0.0
13527	2020-04-09	Yunnan	Mainland China	2020-04-09 23:09:19	184.0	2.0	173.0
13528	2020-04-09	Zhejiang	Mainland China	2020-04-09 23:09:19	1267.0	1.0	1235.0

13529 rows × 7 columns

5. ANALISIS DATA (DATASET CSCEGISandData)

- Mengelompokkan berbagai jenis kasus sesuai tanggal (per 22 Januari - 09 April). 1 hari sebelum pembuatan koding.

In [8]: datewise=covid.groupby(["ObservationDate"]).agg({"Confirmed":'sum',"Recovered":'sum',"Deaths":'sum'})

In [9]: datewise

Out[9]:

ObservationDate	Confirmed	Recovered	Deaths
2020-01-22	555.0	28.0	17.0
2020-01-23	653.0	30.0	18.0
2020-01-24	941.0	36.0	26.0
2020-01-25	1438.0	39.0	42.0
2020-01-26	2118.0	52.0	56.0
...
2020-04-05	1272115.0	260012.0	69374.0
2020-04-06	1345101.0	276515.0	74565.0
2020-04-07	1426096.0	300054.0	81865.0
2020-04-08	1511104.0	328661.0	88338.0
2020-04-09	1595350.0	353975.0	95455.0

79 rows × 3 columns

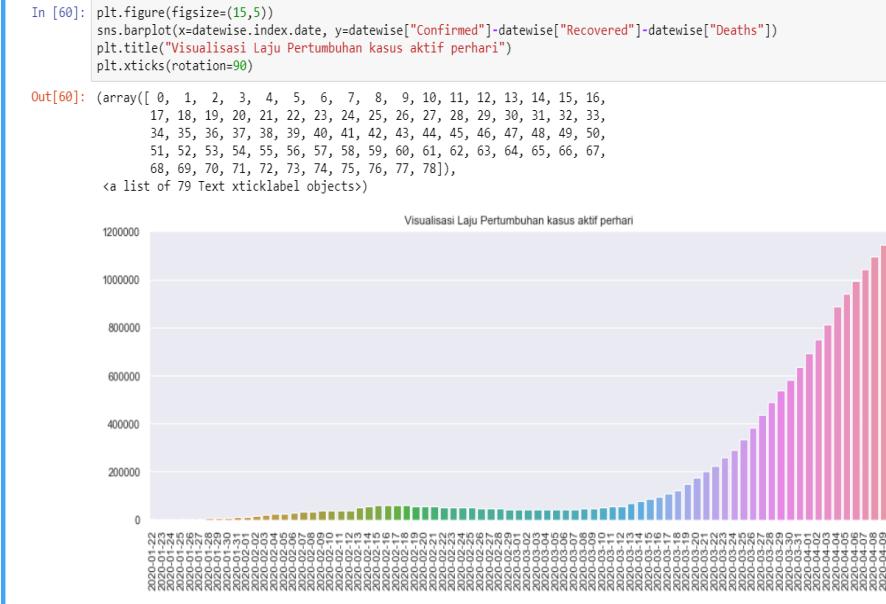
- Print informasi utama

```
In [10]: print("Informasi utama COVID-19")
print("Total negara yang sudah terpapar COVID-19: ",len(covid["Country/Region"].unique()))
print("Total kasus yang sudah dikonfirmasi di dunia: ",datewise["Confirmed"].iloc[-1])
print("Total kasus yang sudah di recovered di dunia: ",datewise["Recovered"].iloc[-1])
print("Total kasus kematian di dunia: ",datewise["Deaths"].iloc[-1])
print("Total kasus aktif di dunia: ",(datewise["Confirmed"].iloc[-1]-datewise["Recovered"].iloc[-1]-datewise["Deaths"].iloc[-1]))
print("Total kasus non-aktif di dunia: ",datewise["Recovered"].iloc[-1]-datewise["Deaths"].iloc[-1])
print("Perkiraan kasus yang di konfirmasi per-hari di dunia: ",np.round(datewise["Confirmed"].iloc[-1]/datewise.shape[0]))
print("Perkiraan kasus yang di recovered per-hari di dunia: ",np.round(datewise["Recovered"].iloc[-1]/datewise.shape[0]))
print("Perkiraan kasus kematian per-hari di dunia: ",np.round(datewise["Deaths"].iloc[-1]/datewise.shape[0]))
print("Perkiraan kasus yang di konfirmasi per-jam di dunia: ",np.round(datewise["Confirmed"].iloc[-1]/((datewise.shape[0])*24)))
print("Perkiraan kasus yang di recovered per-jam di dunia: ",np.round(datewise["Recovered"].iloc[-1]/((datewise.shape[0])*24)))
print("Perkiraan kasus kematian per-jam di dunia: ",np.round(datewise["Deaths"].iloc[-1]/((datewise.shape[0])*24)))
```

- Informasi utama COVID-19
 - Total negara yang sudah terpapar COVID-19: 219
 - Total kasus yang sudah dikonfirmasi di dunia: 1595350.0
 - Total kasus yang sudah di recovered di dunia: 353975.0
 - Total kasus kematian di dunia: 95455.0
 - Total kasus aktif di dunia: 1145920.0
 - Total kasus non-aktif di dunia: 449430.0
 - Perkiraan kasus yang di konfirmasi per-hari di dunia: 20194.0
 - Perkiraan kasus yang di recovered per-hari di dunia: 4481.0
 - Perkiraan kasus kematian per-hari di dunia: 1208.0
 - Perkiraan kasus yang di konfirmasi per-jam di dunia: 841.0

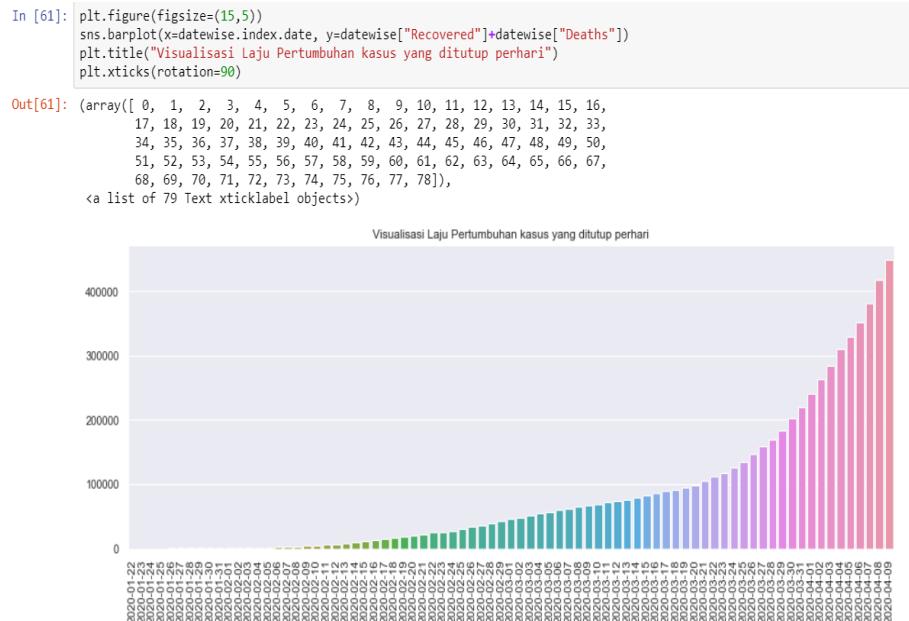
- Perkiraan kasus yang di recovered per-jam di dunia: 187.0
- Perkiraan kasus kematian per-jam di dunia: 50.0

- Ini adalah Grafik Dari Penyebaran Per Hari nya.



- Pada grafik diatas, peningkatan terjadi secara drastis mulai dari tanggal 13 Maret 2020. Ini dikarenakan kasus bertambah secara signifikan namun kasus yang di-recovered dan kasus kematian tidak terlalu banyak. Maka kasus aktif melonjak

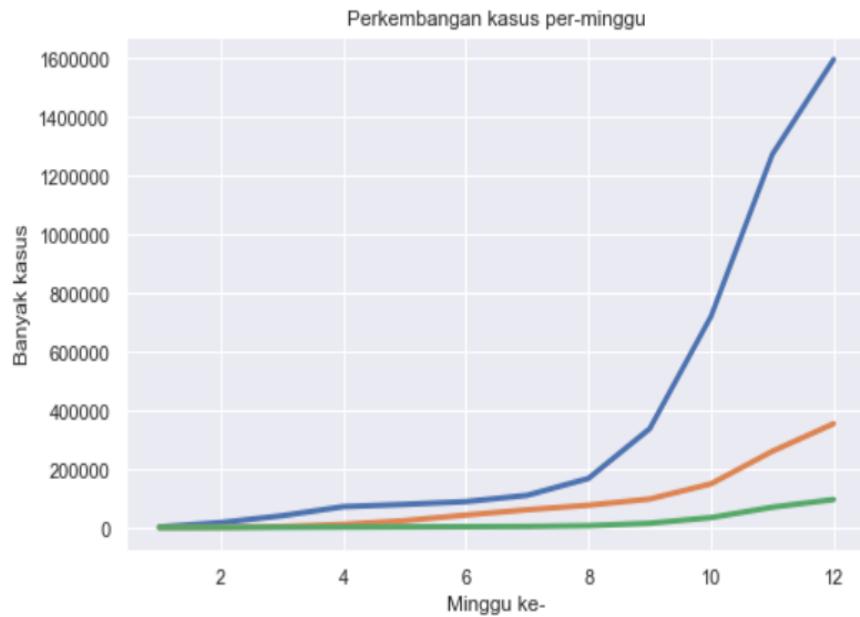
- Ini Adalah Grafik Pertumbuhan Kasus yang ditutup(CLOSED) setiap hari



- Peningkatan jumlah Kasus ditutup pada grafik diatas bisa berarti lebih banyak pasien yang sembuh dari penyakit atau lebih banyak orang meninggal karena COVID-19. Peningkatan terjadi lebih dini dibandingkan dengan Laju Pertumbuhan Kasus aktif pada analisa sebelum ini
- Grafik Kasus Perminggu
- ```
In [79]: datewise["WeekOfYear"] = datewise.index.weekofyear

week_num = []
weekwise_confirmed = []
weekwise_recovered = []
weekwise_deaths = []
w=1
for i in list(datewise["WeekOfYear"].unique()):
 weekwise_confirmed.append(datewise[datewise["WeekOfYear"]==i][["Confirmed"]].iloc[-1])
 weekwise_recovered.append(datewise[datewise["WeekOfYear"]==i][["Recovered"]].iloc[-1])
 weekwise_deaths.append(datewise[datewise["WeekOfYear"]==i][["Deaths"]].iloc[-1])
 week_num.append(w)
 w=w+1

plt.figure(figsize=(8,5))
plt.plot(week_num,weekwise_confirmed,linewidth=3)
plt.plot(week_num,weekwise_recovered,linewidth=3)
plt.plot(week_num,weekwise_deaths,linewidth=3)
plt.ylabel("Banyak kasus")
plt.xlabel("Minggu ke-")
plt.title("Perkembangan kasus per-minggu")
plt.xlabel
```



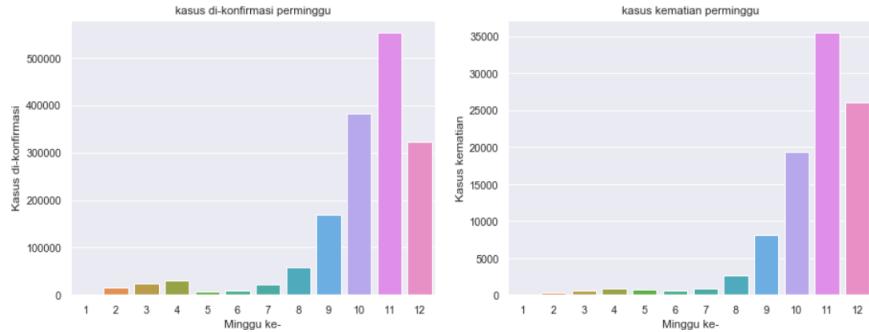
- Keterangan grafik - Perkembangan kasus per-minggu

- kasus di-konfirmasi
- Kasus di-recovered
- Kasus kematian

## - Kasus Kematian Perminggu

```
In [63]: fig, (ax1,ax2) = plt.subplots(1, 2,figsize=(15,5))
sns.barplot(x=week_num,y=pd.Series(weekwise_confirmed).diff().fillna(0),ax=ax1)
sns.barplot(x=week_num,y=pd.Series(weekwise_deaths).diff().fillna(0),ax=ax2)
ax1.set_xlabel("Minggu ke-")
ax2.set_xlabel("Minggu ke-")
ax1.set_ylabel("Kasus di-konfirmasi")
ax2.set_ylabel("Kasus kematian")
ax1.set_title("kasus di-konfirmasi perminggu")
ax2.set_title("kasus kematian perminggu")
```

```
Out[63]: Text(0.5, 1.0, 'kasus kematian perminggu')
```

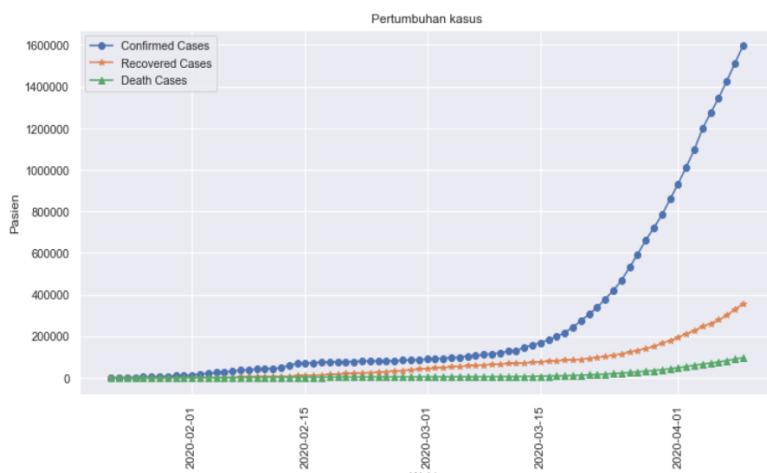


- Pada minggu ke-11, kasus yang di-konfirmasi dan kasus kematian meningkat, namun pada minggu ke-12 grafik melandai.

## - Pertumbuhan Kasus

```
In [64]: plt.figure(figsize=(12,6))
plt.plot(datewise["Confirmed"],marker="o",label="Confirmed Cases")
plt.plot(datewise["Recovered"],marker="*",label="Recovered Cases")
plt.plot(datewise["Deaths"],marker="^",label="Death Cases")
plt.ylabel("Pasien")
plt.xlabel("Waktu")
plt.xticks(rotation=90)
plt.title("Pertumbuhan kasus")
plt.legend()
```

```
Out[64]: <matplotlib.legend.Legend at 0x1b2397059e8>
```



- Laju pertumbuhan kasus yang dikonfirmasi sangat cepat, ini dikarenakan penyebaran virus COVID-19 yang tergolong sangat cepat. April ini sudah hampir 40000 orang yang sudah di-recovered sedangkan untuk kematian hampir mencapai 20000.

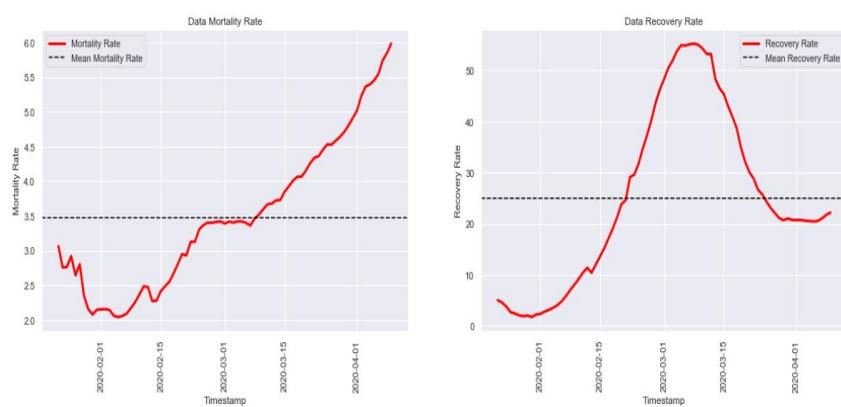
- Mortality dan Recovery Rate analysis di seluruh dunia.
  - Mortality rate adalah tingkat kematian (Kematian/konfirmasi)
  - Recovery rate adalah tingkat kesembuhan (Recovered/konfirmasi)

```
In [65]: #Menghitung Mortality Rate dan Recovery Rate
datewise["Mortality Rate"]=(datewise["Deaths"]/datewise["Confirmed"])*100
datewise["Recovery Rate"]=(datewise["Recovered"]/datewise["Confirmed"])*100
datewise["Active Cases"]=datewise["Confirmed"]-datewise["Recovered"]-datewise["Deaths"]
datewise["Closed Cases"]=datewise["Recovered"]*datewise["Deaths"]

#Plotting Mortality dan Recovery Rate
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(20,6))
ax1.plot(datewise["Mortality Rate"],label='Mortality Rate', linewidth=3,color='red')
ax1.axhline(datewise["Mortality Rate"].mean(),linestyle='--',color='black',label="Mean Mortality Rate")
ax1.set_ylabel("Mortality Rate")
ax1.set_xlabel("Timestamp")
ax1.set_title("Data Mortality Rate")
ax1.legend()
for tick in ax1.get_xticklabels():
 tick.set_rotation(90)
ax2.plot(datewise["Recovery Rate"],label="Recovery Rate", linewidth=3,color='red')
ax2.axhline(datewise["Recovery Rate"].mean(),linestyle='--',color='black',label="Mean Recovery Rate")
ax2.set_ylabel("Recovery Rate")
ax2.set_xlabel("Timestamp")
ax2.set_title("Data Recovery Rate")
ax2.legend()
for tick in ax2.get_xticklabels():
 tick.set_rotation(90)

print("Rata-rata Mortality Rate",datewise["Mortality Rate"].mean())
print("Median Mortality Rate",datewise["Mortality Rate"].median())
print("Rata-rata Recovery Rate",datewise["Recovery Rate"].mean())
print("Median Recovery Rate",datewise["Recovery Rate"].median())

Rata-rata Mortality Rate 3.4806704847250756
Median Mortality Rate 3.4030738233595836
Rata-rata Recovery Rate 25.003797146098126
Median Recovery Rate 21.312515699157863
```



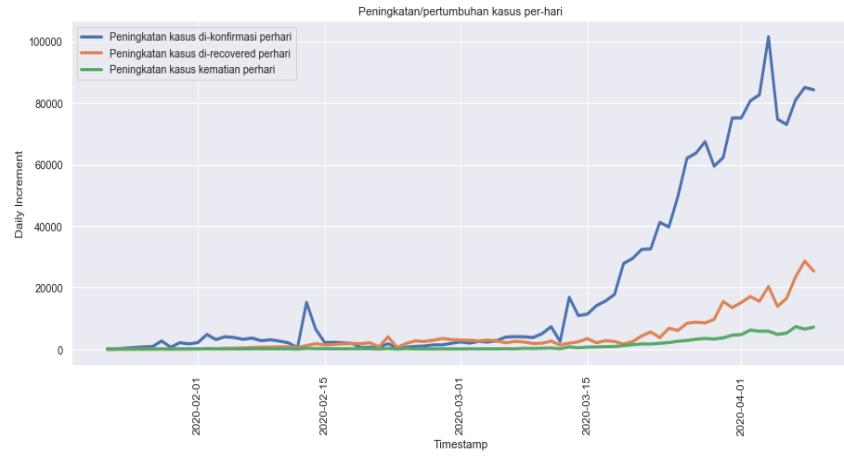
- Grafik mortality rate masih menanjak, sedangkan recovery rate mengalami penurunan yang cukup signifikan. Ini adalah alasan mengapa jumlah kasus aktif meningkat, sedangkan peningkatan kasus yang ditutup erat kaitannya dengan mortality rate. Jelas bahwa mortality rate merupakan indikasi meningkatkan jumlah kasus kematian. Mean untuk mortality rate ada di angka 3.5, butuh sedikit usaha untuk melandaikan grafik ini. Lalu, mean recovery rate berada di angka 25. Dimana grafik sudah mulai menurun ke angka mean.

## - Pertumbuhan Kasus Perhari

```
In [66]: plt.figure(figsize=(15,6))
plt.plot(datewise[\"Confirmed\"].diff().fillna(0),label="Peningkatan kasus di-konfirmasi perhari",linewidth=3)
plt.plot(datewise[\"Recovered\"].diff().fillna(0),label="Peningkatan kasus di-recovered perhari",linewidth=3)
plt.plot(datewise[\"Deaths\"].diff().fillna(0),label="Peningkatan kasus kematian perhari",linewidth=3)
plt.xlabel(\"Timestamp")
plt.ylabel(\"Daily Increment")
plt.title(\"Peningkatan/pertumbuhan kasus per-hari")
plt.xticks(rotation=90)
plt.legend()

print(\"Rata-rata kasus di-konfirmasi: \",np.round(datewise[\"Confirmed\"].diff().fillna(0).mean()))
print(\"Rata-rata kasus di-recovered: \",np.round(datewise[\"Recovered\"].diff().fillna(0).mean()))
print(\"Rata-rata kasus kematian: \",np.round(datewise[\"Deaths\"].diff().fillna(0).mean()))
```

Rata-rata kasus di-konfirmasi: 20187.0  
Rata-rata kasus di-recovered: 4480.0  
Rata-rata kasus kematian: 1208.0

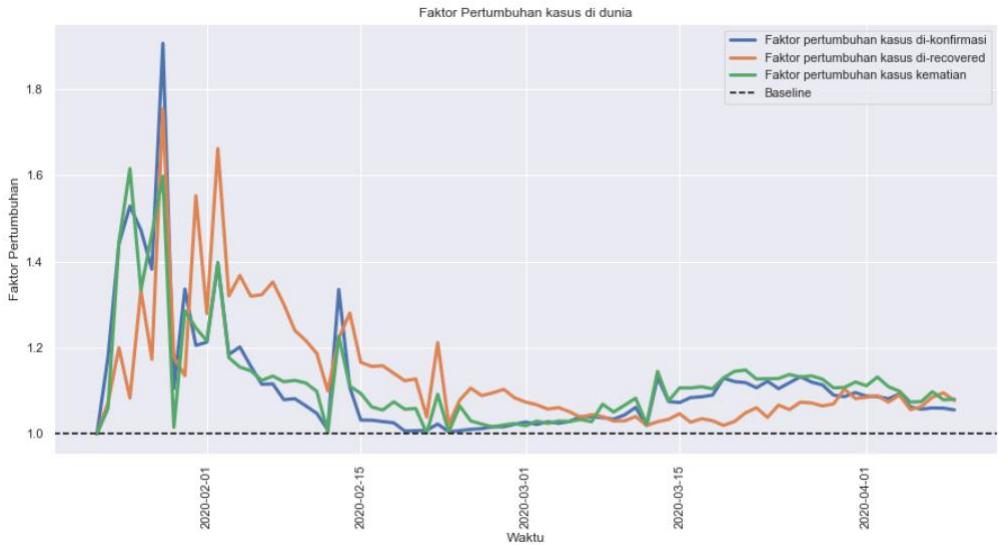


- Faktor pertumbuhan kasus. Untuk menghitung pertumbuhan kasus digunakan perhitungan sebagai berikut = Kasus terbaru/kasus hari sebelumnya. menunjukkan peningkatan kasus correspoding, Grafik menurun menunjukkan tanda positif, sedangkan faktor pertumbuhan konstan di atas adalah tanda pertumbuhan eksponensial, konstan artinya grafik stabil menunjukkan tidak ada perubahan dalam kasus apa pun

## - Pertumbuhan Kasus Diatas Base Line

```
In [67]: daily_increase_confirm=[]
daily_increase_recovered=[]
daily_increase_deaths=[]
for i in range(datewise.shape[0]-1):
 daily_increase_confirm.append(((datewise[\"Confirmed\"].iloc[i+1]/datewise[\"Confirmed\"].iloc[i])))
 daily_increase_recovered.append(((datewise[\"Recovered\"].iloc[i+1]/datewise[\"Recovered\"].iloc[i])))
 daily_increase_deaths.append(((datewise[\"Deaths\"].iloc[i+1]/datewise[\"Deaths\"].iloc[i])))
daily_increase_confirm.insert(0,1)
daily_increase_recovered.insert(0,1)
daily_increase_deaths.insert(0,1)

plt.figure(figsize=(15,7))
plt.plot(datewise.index,daily_increase_confirm,label="Faktor pertumbuhan kasus di-konfirmasi",linewidth=3)
plt.plot(datewise.index,daily_increase_recovered,label="Faktor pertumbuhan kasus di-recovered",linewidth=3)
plt.plot(datewise.index,daily_increase_deaths,label="Faktor pertumbuhan kasus kematian",linewidth=3)
plt.xlabel("Waktu")
plt.ylabel("Faktor Pertumbuhan")
plt.title("Faktor Pertumbuhan kasus di dunia")
plt.axhline(1,linestyle='--',color='black',label="Baseline")
plt.xticks(rotation=90)
plt.legend()
```



- Pertumbuhan kasus yang terus-menerus di atas baseline(1) adalah indikasi yang jelas tentang peningkatan Eksplosif dalam semua bentuk kasus baik yang sudah di-konfirmasi, di-recovered maupun kasus kematian.

### - Analysis By Country (Analisa Negara)

- Menghitung Mortality rate dan Recovery rate per-negara

```
In [68]: countrywise=covid[covid["ObservationDate"]==covid["ObservationDate"].max()].groupby(["Country/Region"]).agg({"Confirmed": "sum", "Recovered": "sum", "Deaths": "sum"}).reset_index()
countrywise["Mortality"]=(countrywise["Deaths"]/countrywise["Confirmed"])*100
countrywise["Recovery"]=(countrywise["Recovered"]/countrywise["Confirmed"])*100
```

Data terbaru negara dengan mortality dan recovery rate

```
In [69]: countrywise
```

```
Out[69]:
```

| Country/Region        | Confirmed | Recovered | Deaths  | Mortality | Recovery  |
|-----------------------|-----------|-----------|---------|-----------|-----------|
| US                    | 461437.0  | 25410.0   | 16478.0 | 3.571018  | 5.506711  |
| Spain                 | 153222.0  | 52165.0   | 15447.0 | 10.081450 | 34.045372 |
| Italy                 | 143626.0  | 28470.0   | 18279.0 | 12.726804 | 19.822316 |
| France                | 118781.0  | 23413.0   | 12228.0 | 10.294576 | 19.711065 |
| Germany               | 118181.0  | 52407.0   | 2607.0  | 2.205938  | 44.344692 |
| ...                   | ...       | ...       | ...     | ...       | ...       |
| Sao Tome and Principe | 4.0       | 0.0       | 0.0     | 0.000000  | 0.000000  |
| Burundi               | 3.0       | 0.0       | 0.0     | 0.000000  | 0.000000  |
| South Sudan           | 3.0       | 0.0       | 0.0     | 0.000000  | 0.000000  |
| Papua New Guinea      | 2.0       | 0.0       | 0.0     | 0.000000  | 0.000000  |
| Timor-Leste           | 1.0       | 0.0       | 0.0     | 0.000000  | 0.000000  |

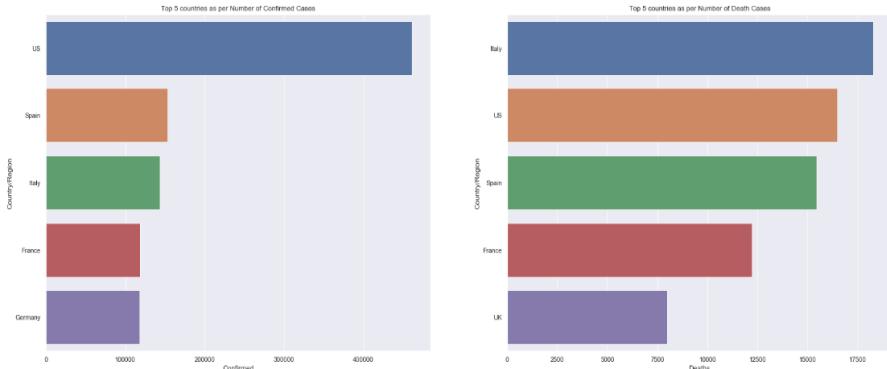
186 rows × 5 columns

- Data terbaru negara dengan mortality dan recovery rate

- 5 negara dengan kasus di-konfirmasi dan kasus kematian terbanyak

```
In [70]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(27,10))
top_5_confirmed=countrywise.sort_values(["Confirmed"],ascending=False).head(5)
top_5_deaths=countrywise.sort_values(["Deaths"],ascending=False).head(5)
sns.barplot(x=top_5_confirmed["Confirmed"],y=top_5_confirmed.index,ax=ax1)
sns.barplot(x=top_5_deaths["Deaths"],y=top_5_deaths.index,ax=ax2)
sns.set(style="darkgrid")
ax1.set_title("Top 5 countries as per Number of Confirmed Cases")
ax2.set_title("Top 5 countries as per Number of Death Cases")
```

Out[70]: Text(0.5, 1.0, 'Top 5 countries as per Number of Death Cases')

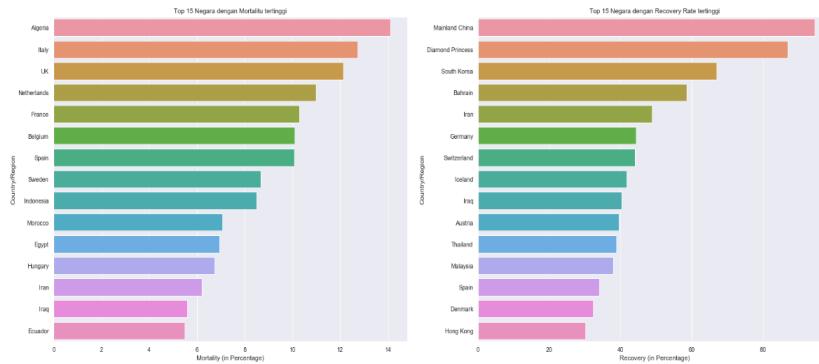


- Faktor yang mungkin menyebabkan negara diatas menjadi 5 negara paling terkena dampak COVID-19 adalah pada link berikut (1)Tourist Data: <https://worldpopulationreview.com/countries/most-visited-countries/> (2)International Students Data: <https://www.easyuni.com/advice/top-countries-with-most-international-students-1184/>. Jika kita memeriksa daftar negara yang sesuai dengan jumlah wisatawan yang berkunjung dari tautan yang disebutkan di atas, negara-negara utama adalah Perancis, Spanyol, AS, Cina, Italia, Meksiko, Inggris, Turki, Jerman, Thailand. Hal lain yang perlu dipertimbangkan sebagian besar negara yang disebutkan di atas juga memiliki jumlah Siswa Internasional terbanyak. Semua dari mereka adalah negara yang paling terkena dampak karena COVID-19.

- 15 Negara teratas berdasarkan Mortatly Rate dan Recovery Rate dengan lebih dari 500 kasus di-konfirmasi.

```
In [71]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(27,10))
countrywise_plot_mortal=countrywise[countrywise["Confirmed"]>500].sort_values([["Mortality"]],ascending=False).head(15)
sns.barplot(x=countrywise_plot_mortal["Mortality"],y=countrywise_plot_mortal.index,ax=ax1)
ax1.set_title("Top 15 Negara dengan Mortalitudo tertinggi")
ax1.set_xlabel("Mortality (in Percentage)")
countrywise_plot_recover=countrywise[countrywise["Confirmed"]>500].sort_values([["Recovery"]],ascending=False).head(15)
sns.barplot(x=countrywise_plot_recover["Recovery"],y=countrywise_plot_recover.index, ax=ax2)
ax2.set_title("Top 15 Negara dengan Recovery Rate tertinggi")
ax2.set_xlabel("Recovery (in Percentage)")
```

Out[71]: Text(0.5, 0, 'Recovery (in Percentage)')



```
In [72]: no_recovered_countries=countrywise[(countrywise["Confirmed"]>50)&(countrywise["Recovered"]==0)][["Confirmed","Deaths"]]
no_recovered_countries["Mortality Rate"]=(no_recovered_countries["Deaths"]/no_recovered_countries["Confirmed"])*100
no_recovered_countries[no_recovered_countries["Mortality Rate"]>0].sort_values(["Mortality Rate"],ascending=False)
```

Out[72]:

| Country/Region | Confirmed | Deaths | Mortality Rate |
|----------------|-----------|--------|----------------|
| Serbia         | 2867.0    | 66.0   | 2.302058       |

- Serbia adalah negara dengan Mortality rate yang cukup tinggi karena jumlah kasus di-konfirmasi jauh di atas 1000 dengan jumlah Mortality rate yang cukup banyak tanpa pasien yang Sembuh.

- Negara-negara dengan lebih dari 100 Kasus di-konfirmasi dan Tidak Ada Kematian dengan Recovery rate yang sangat tinggi.

```
In [73]: no_deaths=countrywise[(countrywise["Confirmed"]>100)&(countrywise["Deaths"]==0)][["Confirmed","Deaths"]]
no_deaths=no_deaths[no_deaths["Recovered"]>0].sort_values(["Recovered"],ascending=False).drop(["Mortality"],1)
```

Out[73]:

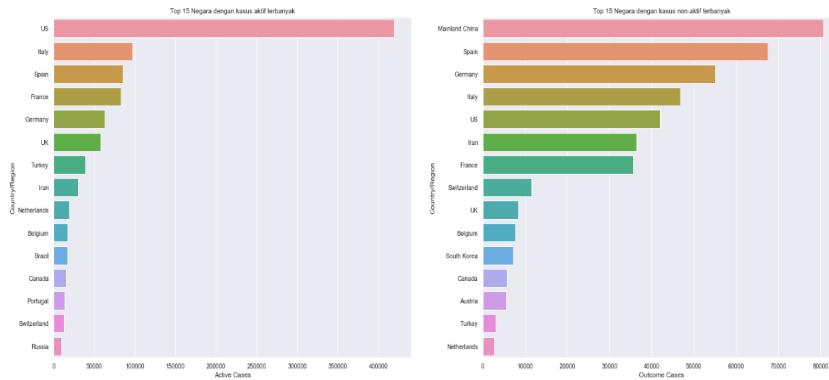
| Country/Region | Confirmed | Recovered | Deaths | Recovery  |
|----------------|-----------|-----------|--------|-----------|
| Cambodia       | 119.0     | 62.0      | 0.0    | 52.100840 |
| Vietnam        | 265.0     | 128.0     | 0.0    | 50.196078 |
| Djibouti       | 135.0     | 25.0      | 0.0    | 18.518519 |
| Rwanda         | 110.0     | 7.0       | 0.0    | 6.363636  |
| Guinea         | 194.0     | 11.0      | 0.0    | 5.670103  |

- Vietnam mampu menghadapi COVID-19 dengan cukup baik tanpa Kematian yang tercatat sejauh ini dengan Recovery rate yang cukup sehat. Vietnam adalah negara pertama yang memberi tahu Organisasi Kesehatan Dunia tentang Transmisi Manusia ke Manusia dari COVID-19.

- Top 15 Negara dengan kasus non-aktif terbanyak

```
In [74]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(27,10))
countrywise[["Active Cases"]]=(countrywise[["Confirmed"]]-countrywise[["Recovered"]]-countrywise[["Deaths"]])
countrywise[["Outcome Cases"]]=(countrywise[["Recovered"]]+countrywise[["Deaths"]])
top_15_active=countrywise.sort_values(["Active Cases"],ascending=False).head(15)
top_15_outcome=countrywise.sort_values(["Outcome Cases"],ascending=False).head(15)
sns.barplot(x=top_15_active["Active Cases"],y=top_15_active.index,ax=ax1)
sns.barplot(x=top_15_outcome["Outcome Cases"],y=top_15_outcome.index,ax=ax2)
ax1.set_title("Top 15 Negara dengan kasus aktif terbanyak")
ax2.set_title("Top 15 Negara dengan kasus non-aktif terbanyak")
```

Out[74]: Text(0.5, 1.0, 'Top 15 Negara dengan kasus non-aktif terbanyak')

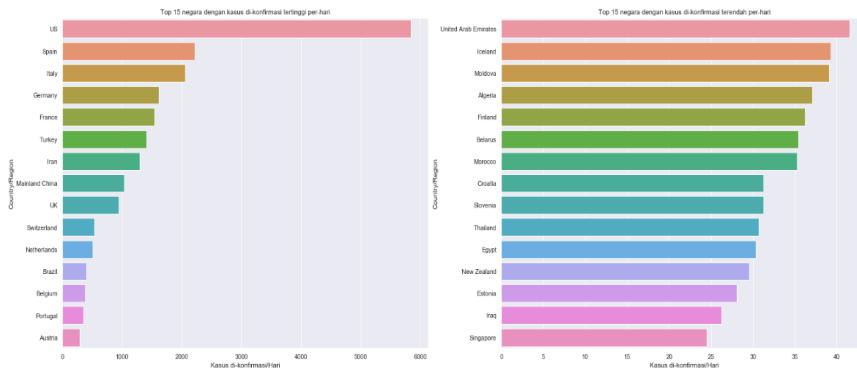


```
In [26]: country_date=covid.groupby(["Country/Region","ObservationDate"]).agg({"Confirmed":'sum','Recovered':'sum','Deaths':'sum'})
confirm_rate=[]
for country in countrywise.index:
 days=country_date.ix[country].shape[0]
 confirm_rate.append((countrywise.ix[country][["Confirmed"]]/days))
countrywise[["Kasus di-konfirmasi/Hari"]]=confirm_rate
```

- Negara-negara dengan kasus di-konfirmasi tertinggi dan terendah per-hari

```
In [27]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(27,10))
top_15_ccpd=countrywise.sort_values(["Kasus di-konfirmasi/Hari"],ascending=False).head(15)
sns.barplot(y=top_15_ccpd.index,x=top_15_ccpd["Kasus di-konfirmasi/Hari"],ax=ax1)
ax1.set_title("Top 15 negara dengan kasus di-konfirmasi tertinggi per hari")
bottom_15_ccpd=countrywise[["Confirmed"]].tail(1000).sort_values(["Kasus di-konfirmasi/Hari"],ascending=False).tail(15)
sns.barplot(y=bottom_15_ccpd.index,x=bottom_15_ccpd["Kasus di-konfirmasi/Hari"],ax=ax2)
ax2.set_title("Top 15 negara dengan kasus di-konfirmasi terendah per hari")
```

Out[27]: Text(0.5, 1.0, 'Top 15 negara dengan kasus di-konfirmasi terendah per hari')



- Cina Daratan telah mencatat jumlah Closed cases tertinggi karena Recovery rate mereka mengejutkan mencatat 85% +, alasan mengapa Italia telah menempati peringkat kedua di antara negara-negara dengan jumlah Closed cases tertinggi adalah karena jumlah Kematian di Italia, Death rate di Italia adalah 10% + di mana COVID-19 memiliki Death rate 2-3%. Confirmed cases/day yang Dikonfirmasi adalah indikasi yang jelas mengapa US memiliki jumlah Active cases

tertinggi saat ini. Ada 5000+ kasus per hari. Peningkatan terus terjadi setiap hari.

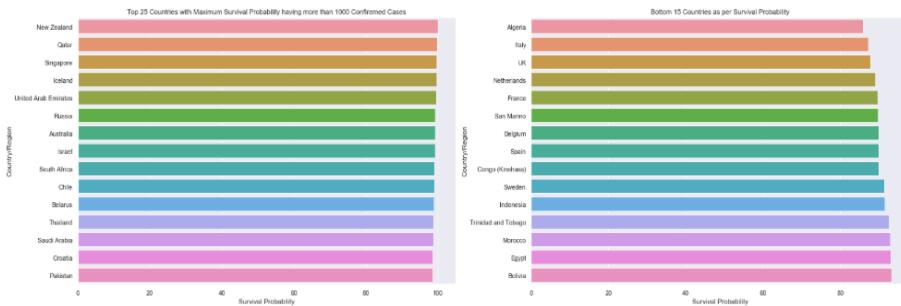
### - Survival probability

```
In [28]: fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(27,8))
countrywise["Survival Probability"]=(1-(countrywise["Deaths"]/countrywise["Confirmed"]))*100
top_25_survival=countrywise[countrywise["Confirmed"]>1000].sort_values(["Survival Probability"],ascending=False).head(15)
sns.barplot(x=top_25_survival["Survival Probability"],y=top_25_survival.index,ax=ax1)
ax1.set_title("Top 25 Countries with Maximum Survival Probability having more than 1000 Confirmed Cases")
print('Mean Survival Probability across all countries',countrywise["Survival Probability"].mean())
print('Median Survival Probability across all countries',countrywise["Survival Probability"].median())
print('Mean Death Probability across all countries',100-countrywise["Survival Probability"].mean())
print('Median Death Probability across all countries',100-countrywise["Survival Probability"].median())

Bottom_5_countries=countrywise[countrywise["Confirmed"]<100].sort_values(["Survival Probability"],ascending=True).head(15)
sns.barplot(x=Bottom_5_countries["Survival Probability"],y=Bottom_5_countries.index,ax=ax2)
plt.title("Bottom 15 Countries as per Survival Probability")

Mean Survival Probability across all countries 96.0182977269845
Median Survival Probability across all countries 97.73873333803513
Mean Death Probability across all countries 3.981702273015472
Median Death Probability across all countries 2.26126661964868
```

Out[28]: Text(0.5, 1.0, 'Bottom 15 Countries as per Survival Probability')



- Survival Probability satu-satunya grafik yang terlihat paling menjanjikan! Memiliki probabilitas kelangsungan hidup rata-rata 95% + di semua negara tetapi turun sedikit setiap hari. Perbedaan antara Mean dan Death Probability adalah indikasi yang jelas bahwa ada beberapa negara dengan tingkat kematian yang sangat tinggi mis. Italia, Aljazair, Inggris, dll.. .

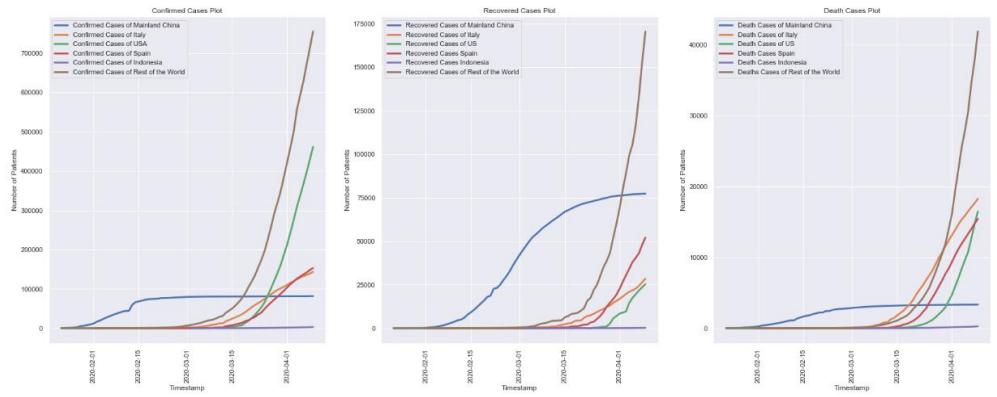
## - Perbandingan dari China, Italy, US, Spain, Indonesia dan Seluruh Dunia

```
In [29]: china_data=covid[covid["Country/Region"]=="Mainland China"]
Italy_data=covid[covid["Country/Region"]=="Italy"]
US_data=covid[covid["Country/Region"]=="US"]
spain_data=covid[covid["Country/Region"]=="Spain"]
indonesia_data=covid[covid["Country/Region"]=="Indonesia"]
rest_of_world=covid[(covid["Country/Region"]!="Mainland China")&(covid["Country/Region"]!="Italy")&(covid["Country/Region"]!="US")]

datewise_china=china_data.groupby(["ObservationDate"]).agg({"Confirmed":'sum','Recovered':'sum','Deaths':'sum'})
datewise_Italy=Italy_data.groupby(["ObservationDate"]).agg({"Confirmed":'sum','Recovered':'sum','Deaths':'sum'})
datewise_US=US_data.groupby(["ObservationDate"]).agg({"Confirmed":'sum','Recovered':'sum','Deaths':'sum'})
datewise_Spain=spain_data.groupby(["ObservationDate"]).agg({"Confirmed":'sum','Recovered':'sum','Deaths':'sum'})
datewise_Indo=indonesia_data.groupby(["ObservationDate"]).agg({"Confirmed":'sum','Recovered':'sum','Deaths':'sum"})
datewise_restofworld=rest_of_world.groupby(["ObservationDate"]).agg({"Confirmed":'sum','Recovered':'sum','Deaths':'sum'})
```

```
In [30]: fig, (ax1, ax2, ax3) = plt.subplots(1, 3,figsize=(28,10))
ax1.plot(datewise_china["Confirmed"],label="Confirmed Cases of Mainland China",linewidth=3)
ax1.plot(datewise_Italy["Confirmed"],label="Confirmed Cases of Italy",linewidth=3)
ax1.plot(datewise_US["Confirmed"],label="Confirmed Cases of USA",linewidth=3)
ax1.plot(datewise_Spain["Confirmed"],label="Confirmed Cases of Spain",linewidth=3)
ax1.plot(datewise_Indo["Confirmed"],label="Confirmed Cases of Indonesia",linewidth=3)
ax1.plot(datewise_restofworld["Confirmed"],label="Confirmed Cases of Rest of the World",linewidth=3)
ax1.set_title("Confirmed Cases Plot")
ax1.set_ylabel("Number of Patients")
ax1.set_xlabel("Timestamp")
ax1.legend()
for tick in ax1.get_xticklabels():
 tick.set_rotation(90)
ax2.plot(datewise_china["Recovered"],label="Recovered Cases of Mainland China",linewidth=3)
ax2.plot(datewise_Italy["Recovered"],label="Recovered Cases of Italy",linewidth=3)
ax2.plot(datewise_US["Recovered"],label="Recovered Cases of US",linewidth=3)
ax2.plot(datewise_Spain["Recovered"],label="Recovered Cases Spain",linewidth=3)
ax2.plot(datewise_Indo["Recovered"],label="Recovered Cases Indonesia",linewidth=3)
ax2.plot(datewise_restofworld["Recovered"],label="Recovered Cases of Rest of the World",linewidth=3)
ax2.set_title("Recovered Cases Plot")
ax2.set_ylabel("Number of Patients")
ax2.set_xlabel("Timestamp")
ax2.legend()
for tick in ax2.get_xticklabels():
 tick.set_rotation(90)
ax3.plot(datewise_china["Deaths"],label='Death Cases of Mainland China',linewidth=3)
ax3.plot(datewise_Italy["Deaths"],label='Death Cases of Italy',linewidth=3)
ax3.plot(datewise_US["Deaths"],label='Death Cases of US',linewidth=3)
ax3.plot(datewise_Spain["Deaths"],label='Death Cases Spain',linewidth=3)
ax3.plot(datewise_Indo["Deaths"],label='Death Cases Indonesia',linewidth=3)
ax3.plot(datewise_restofworld["Deaths"],label="Deaths Cases of Rest of the World",linewidth=3)
ax3.set_title("Death Cases Plot")
ax3.set_ylabel("Number of Patients")
ax3.set_xlabel("Timestamp")
ax3.legend()
for tick in ax3.get_xticklabels():
 tick.set_rotation(90)
```

- Dari codingan diatas didapat grafik sebagai berikut :

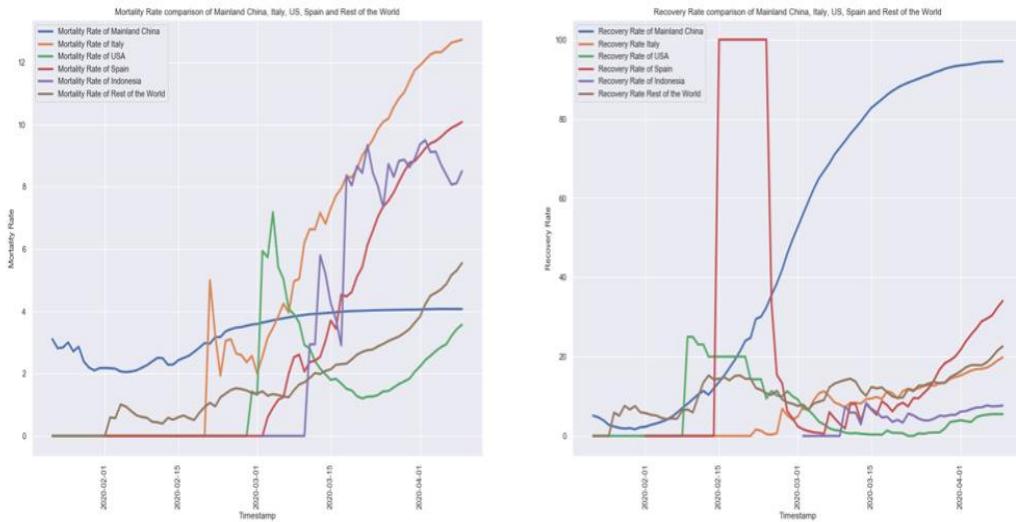


- Keterangan pada grafik diatas, Perbandingan plot grafik antara kasus yang terkonfirmasi, kasus pemulihan dan kasus kematian
  - o Cina telah mampu melakukan “flatten the curve” dengan melihat grafik Confirmed cases dan Death cases mereka. Dengan recovery rate yang mengejutkan. Dilihat bahwa jumlah pasien di china yang kasus terkonfirmasi ada kurang dari 100.000 pasien, untuk kasus kematian, China memiliki pasien yang lebih kurang dari 10.000 pasien.
  - o US tampaknya memiliki kontrol yang baik pada Kematian, tetapi jumlah orang yang terkena dampak akan keluar dari yang ditangani.

```
In [31]: datewise_china["Mortality"]=(datewise_china["Deaths"]/datewise_china["Confirmed"])*100
datewise_Italy["Mortality"]=(datewise_Italy["Deaths"]/datewise_Italy["Confirmed"])*100
datewise_US["Mortality"]=(datewise_US["Deaths"]/datewise_US["Confirmed"])*100
datewise_Spain["Mortality"]=(datewise_Spain["Deaths"]/datewise_Spain["Confirmed"])*100
datewise_Indo["Mortality"]=(datewise_Indo["Deaths"]/datewise_Indo["Confirmed"])*100
datewise_restofworld["Mortality"]=(datewise_restofworld["Deaths"]/datewise_restofworld["Confirmed"])*100

datewise_china["Recovery"]=(datewise_china["Recovered"]/datewise_china["Confirmed"])*100
datewise_Italy["Recovery"]=(datewise_Italy["Recovered"]/datewise_Italy["Confirmed"])*100
datewise_US["Recovery"]=(datewise_US["Recovered"]/datewise_US["Confirmed"])*100
datewise_Spain["Recovery"]=(datewise_Spain["Recovered"]/datewise_Spain["Confirmed"])*100
datewise_Indo["Recovery"]=(datewise_Indo["Recovered"]/datewise_Indo["Confirmed"])*100
datewise_restofworld["Recovery"]=(datewise_restofworld["Recovered"]/datewise_restofworld["Confirmed"])*100
```

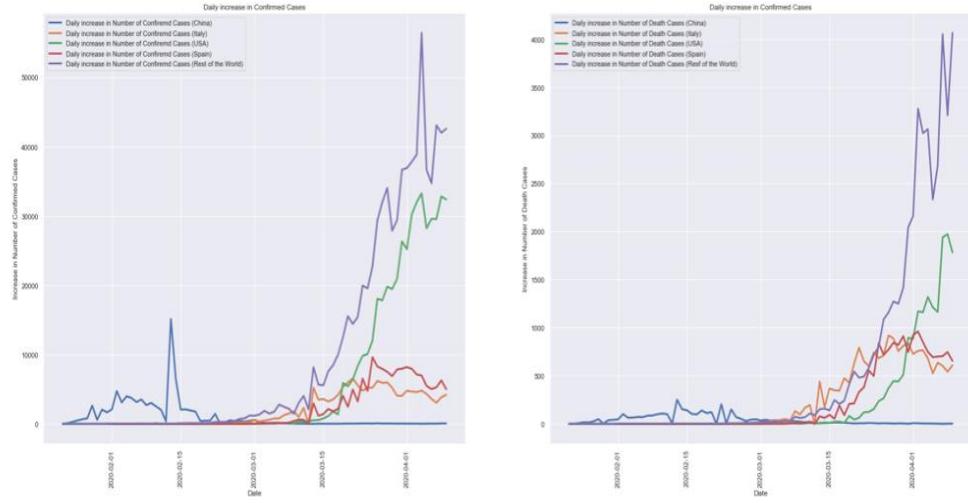
```
In [32]: fig, (ax1,ax2) = plt.subplots(1, 2,figsize=(28,10))
ax1.plot(datewise_china["Mortality"],label="Mortality Rate of Mainland China",linewidth=3)
ax1.plot(datewise_Italy["Mortality"],label="Mortality Rate of Italy",linewidth=3)
ax1.plot(datewise_US["Mortality"],label="Mortality Rate of USA",linewidth=3)
ax1.plot(datewise_Spain["Mortality"],label="Mortality Rate of Spain",linewidth=3)
ax1.plot(datewise_Indo["Mortality"],label="Mortality Rate of Indonesia",linewidth=3)
ax1.plot(datewise_restofworld["Mortality"],label="Mortality Rate of Rest of the World",linewidth=3)
ax1.set_xlabel("Timestamp")
ax1.set_title("Mortality Rate comparison of Mainland China, Italy, US, Spain and Rest of the World")
ax1.legend()
for tick in ax1.get_xticklabels():
 tick.set_rotation(90)
ax2.plot(datewise_china["Recovery"],label="Recovery Rate of Mainland China",linewidth=3)
ax2.plot(datewise_Italy["Recovery"],label="Recovery Rate Italy",linewidth=3)
ax2.plot(datewise_US["Recovery"],label="Recovery Rate of USA",linewidth=3)
ax2.plot(datewise_Spain["Recovery"],label="Recovery Rate of Spain",linewidth=3)
ax2.plot(datewise_Indo["Recovery"],label="Recovery Rate of Indonesia",linewidth=3)
ax2.plot(datewise_restofworld["Recovery"],label="Recovery Rate Rest of the World",linewidth=3)
ax2.set_xlabel("Timestamp")
ax2.set_title("Recovery Rate comparison of Mainland China, Italy, US, Spain and Rest of the World")
ax2.legend()
for tick in ax2.get_xticklabels():
 tick.set_rotation(90)
```



- Keterangan pada Grafik diatas, Perbandingan grafik antara Mortality rate dan Recovery rate
- Melepas Recovery rate Spanyol adalah pertanda baik tetapi tidak ada apa-apanya dibandingkan dengan Death rate. Ini adalah pertanda yang mengkhawatirkan bagi US karena Recovery rate menurun dengan Death rate mulai berkurang

```
In [33]: fig, (ax1,ax2) = plt.subplots(1, 2,figsize=(30,12))
ax1.plot(datewise_china["Confirmed"].diff().fillna(0),label='Daily increase in Number of Confirmed Cases (China)',linewidth=3)
ax1.plot(datewise_Italy["Confirmed"].diff().fillna(0),label='Daily increase in Number of Confirmed Cases (Italy)',linewidth=3)
ax1.plot(datewise_US["Confirmed"].diff().fillna(0),label='Daily increase in Number of Confirmed Cases (USA)',linewidth=3)
ax1.plot(datewise_Spain["Confirmed"].diff().fillna(0),label='Daily increase in Number of Confirmed Cases (Spain)',linewidth=3)
ax1.plot(datewise_restofworld["Confirmed"].diff().fillna(0),label='Daily increase in Number of Confirmed Cases (Rest of the World)',linewidth=3)
ax1.set_xlabel('Date')
ax1.set_ylabel("Increase in Number of Confirmed Cases")
ax1.set_title("Daily increase in Confirmed Cases")
ax1.legend()
for tick in ax1.get_xticklabels():
 tick.set_rotation(90)
ax2.plot(datewise_china["Deaths"].diff().fillna(0),label='Daily increase in Number of Death Cases (China)',linewidth=3)
ax2.plot(datewise_Italy["Deaths"].diff().fillna(0),label='Daily increase in Number of Death Cases (Italy)',linewidth=3)
ax2.plot(datewise_US["Deaths"].diff().fillna(0),label='Daily increase in Number of Death Cases (USA)',linewidth=3)
ax2.plot(datewise_Spain["Deaths"].diff().fillna(0),label='Daily increase in Number of Death Cases (Spain)',linewidth=3)
ax2.plot(datewise_restofworld["Deaths"].diff().fillna(0),label='Daily increase in Number of Death Cases (Rest of the World)',linewidth=3)
ax2.set_xlabel('Date')
ax2.set_ylabel("Increase in Number of Death Cases")
ax2.set_title("Daily increase in Death Cases")
ax2.legend()
for tick in ax2.get_xticklabels():
 tick.set_rotation(90)
```

- Didapat grafik sebagai berikut :



- Keterangan pada Grafik diatas, Perbandingan grafik harian untuk kasus yang terkonfirmasi dengan kasus kematian
- Penurun grafik pada Spanyol dan Italy adalah peningkatkan harian dalam jumlah kasus yang terkonfirmasi, dan penurunan kasus kematian harian. ini adalah tanda yang sedikit positif.

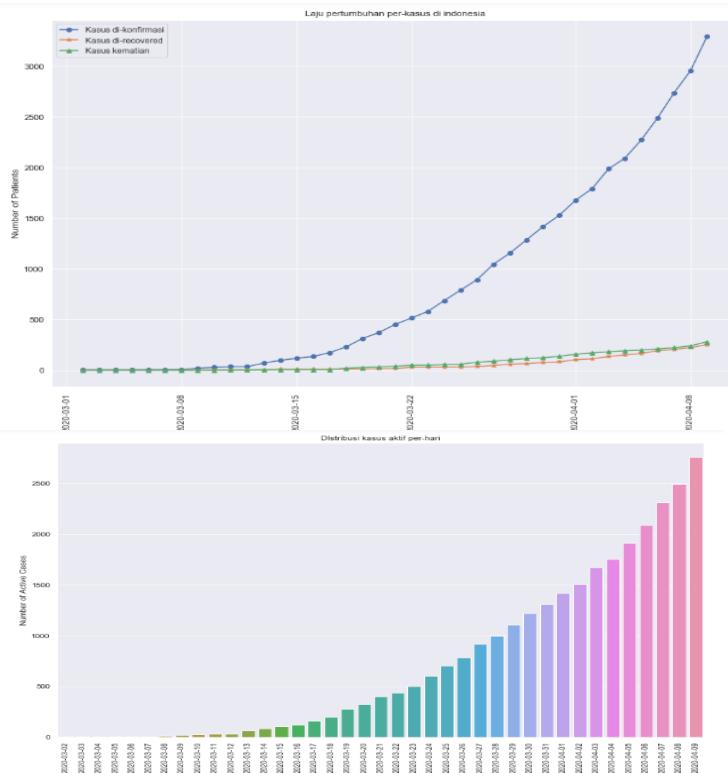
## - Analisa Data COVID-19 Indonesia (DATASET CSCEGISandData)

- Data COVID-19 Indonesia per-tanggal 09 April 2020.

```
In [34]: indonesia_data[covid["Country/Region"]=="Indonesia"]
datewise_indo=indonesia_data.groupby(["ObservationDate"]).agg({"Confirmed":'sum', "Recovered":'sum', "Deaths":'sum'})
print(datewise_indo.iloc[-1])
print("Total Kasus Aktif: ",datewise_indo["Confirmed"].iloc[-1]-datewise_indo["Recovered"].iloc[-1]-datewise_indo["Deaths"].iloc[-1])
print("Total Kasus Non-Aktif: ",datewise_indo["Recovered"].iloc[-1]+datewise_indo["Deaths"].iloc[-1])
<ipython-input-34-1f3a2a2a2a2a>
Confirmed 3293.0
Recovered 252.0
Deaths 280.0
Name: 2020-04-09 00:00:00, dtype: float64
Total Kasus Aktif: 2761.0
Total Kasus Non-Aktif: 532.0
```

- Kasus dikonfirmasi pertanggal 09 April 2020 ada 3293, di-recovered 252 dan kasus kematian mencapai 280. Kasus aktif sebanyak 2761 dan kasus non-aktif 532.

```
In [35]: fig, (ax1,ax2) = plt.subplots(2, 1,figsize=(15,22))
ax1.plot(datewise_indo["Confirmed"],marker='o',label="Kasus di-konfirmasi")
ax1.plot(datewise_indo["Recovered"],marker='*',label="Kasus di-recovered")
ax1.plot(datewise_indo["Deaths"],marker='^',label="Kasus kematian")
ax1.set_ylabel("Number of Patients")
ax1.set_xlabel("Date")
ax1.legend()
ax1.set_title("Laju pertumbuhan per-kasus di indonesia")
for tick in ax1.get_xticklabels():
 tick.set_rotation(90)
sns.barplot(datewise_indo.index.date,datewise_indo["Confirmed"]-datewise_indo["Recovered"]-datewise_indo["Deaths"],ax=ax2)
ax2.set_xlabel("Date")
ax2.set_ylabel("Number of Active Cases")
ax2.set_title("Distribusi kasus aktif per-hari")
for tick in ax2.get_xticklabels():
 tick.set_rotation(90)
```



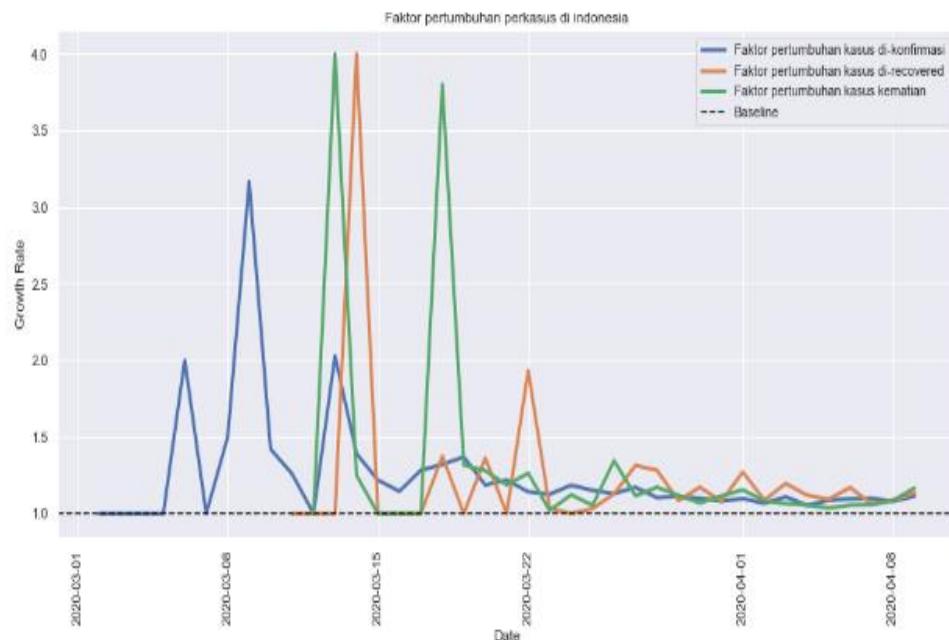
- Keterangan pada Grafik diatas, Laju Pertumbuhan dan Distribusi perhari
- Grafik laju pertumbuhan kasus di indonesia, Grafik ke 2 menunjukan distribusi angka kasus aktif yang terus melonjak per-hari.

```
In [36]: indo_increase_confirm=[]
indo_increase_recover=[]
indo_increase_deaths=[]
for i in range(datewise_indo.shape[0]-1):
 indo_increase_confirm.append(((datewise_indo["Confirmed"].iloc[i+1])/datewise_indo["Confirmed"].iloc[i]))
 indo_increase_recover.append(((datewise_indo["Recovered"].iloc[i+1])/datewise_indo["Recovered"].iloc[i]))
 indo_increase_deaths.append(((datewise_indo["Deaths"].iloc[i+1])/datewise_indo["Deaths"].iloc[i]))
indo_increase_confirm.insert(0,1)
indo_increase_recover.insert(0,1)
indo_increase_deaths.insert(0,1)

plt.figure(figsize=(15,7))
plt.plot(datewise_indo.index,indo_increase_confirm,label="Faktor pertumbuhan kasus di-konfirmasi",linewidth=3)
plt.plot(datewise_indo.index,indo_increase_recover,label="Faktor pertumbuhan kasus di-recovered",linewidth=3)
plt.plot(datewise_indo.index,indo_increase_deaths,label="Faktor pertumbuhan kasus kematian",linewidth=3)
plt.axhline(1,linestyle='--',color="black",label="Baseline")
plt.xticks(rotation=90)
plt.title("Faktor pertumbuhan perkasus di indonesia")
plt.ylabel("Growth Rate")
plt.xlabel("Date")
plt.legend()
```

Out[36]: <matplotlib.legend.Legend at 0x1b23a144fd0>

- Didapat Grafik sebagai berikut



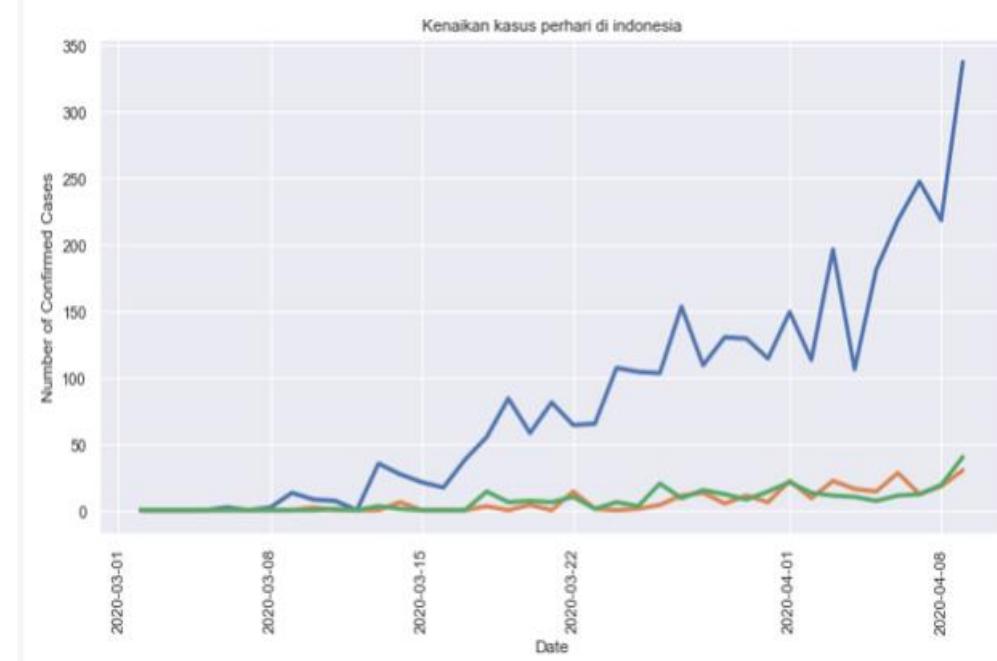
- Faktor pertumbuhan perkasus di indonesia dan garis baseline
- Terlihat kemajuan pada beberapa hari terakhir, walaupun angka belum mendekati baseline(1)

```
In [37]: plt.figure(figsize=(12,6))
plt.plot(datewise_indo["Confirmed"].diff().fillna(0),label="Daily increase in Co
plt.plot(datewise_indo["Recovered"].diff().fillna(0),label="Daily increase in Re
plt.plot(datewise_indo["Deaths"].diff().fillna(0),label="Daily increase in Death
plt.ylabel("Number of Confirmed Cases")
plt.xlabel("Date")
plt.title("Kenaikan kasus perhari di indonesia")
plt.xticks(rotation=90)

print("Rata-rata kenaikan angka kasus di-konfirmasi setiap hari: ",np.round(date
print("Rata-rata kenaikan angka kasus di-recovered setiap hari: ",np.round(datew
print("Rata-rata kenaikan angka kasus kematian setiap hari: ",np.round(datewise_
```

Rata-rata kenaikan angka kasus di-konfirmasi setiap hari: 84.0  
Rata-rata kenaikan angka kasus di-recovered setiap hari: 6.0  
Rata-rata kenaikan angka kasus kematian setiap hari: 7.0

- Didapat Grafik sebagai berikut



- Keterangan pada Grafik diatas, Kenaikan kasus perhari di Indonesia
- Rata-rata kasus baru perhari mencapai 84 kasus, 6 kasus berhasil di-recovered dan 7 kasus kematian.

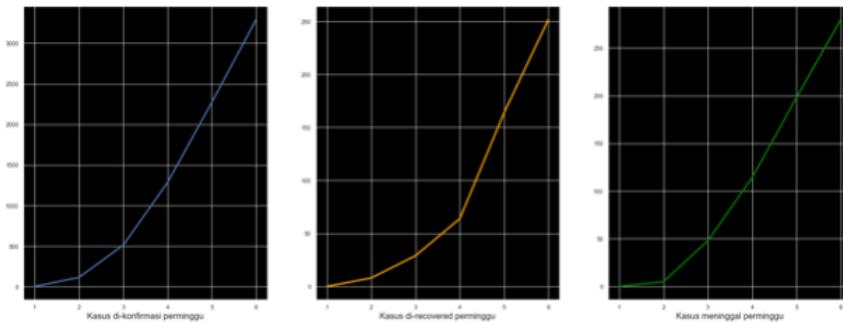
```
In [38]: fig, (ax1, ax2, ax3) = plt.subplots(1, 3, figsize=(28,10))
datewise_indo["WeekOfYear"] = datewise_indo.index.weekofyear

week_num_indo = []
indo_weekwise_confirmed = []
indo_weekwise_recovered = []
indo_weekwise_deaths = []
w=1
for i in list(datewise_indo["WeekOfYear"].unique()):
 indo_weekwise_confirmed.append(datewise_indo[datewise_indo["WeekOfYear"]==i])
 indo_weekwise_recovered.append(datewise_indo[datewise_indo["WeekOfYear"]==i])
 indo_weekwise_deaths.append(datewise_indo[datewise_indo["WeekOfYear"]==i])
 week_num_indo.append(w)
 w=w+1

ax1.plot(week_num_indo,indo_weekwise_confirmed,linewidth=3)
ax1.set_facecolor('black')
ax1.set_xlabel('Kasus di-konfirmasi per minggu', fontsize=16)
ax2.plot(week_num_indo,indo_weekwise_recovered,linewidth=3,color='orange')
ax2.set_facecolor('black')
ax2.set_xlabel('Kasus di-recovered per minggu', fontsize=16)

No handles with labels found to put in legend.
```

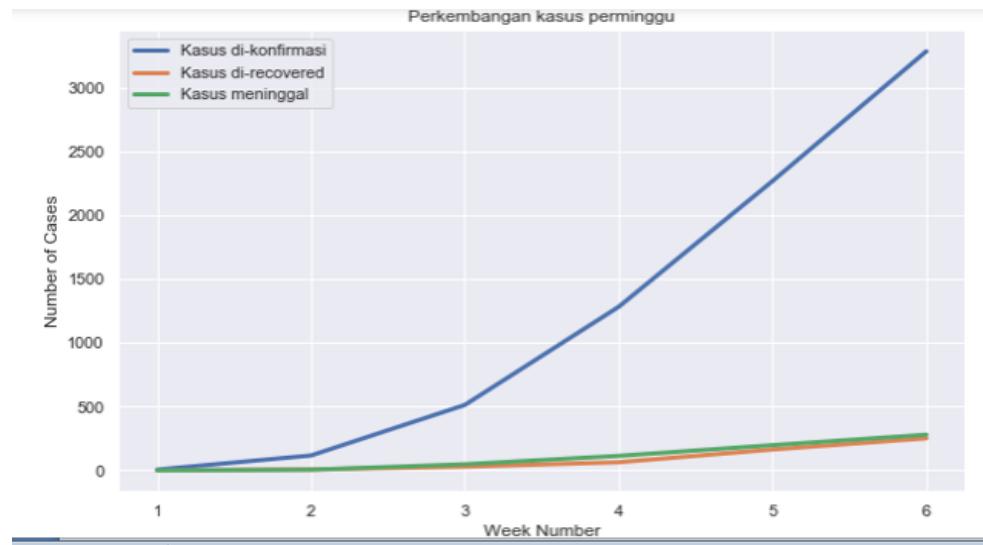
Out[38]: <matplotlib.legend.Legend at 0x1b23a268e10>



- Keterangan pada Grafik diatas, Kasus perminggu
  - o Pada minggu ke-6 kasus di-konfirmasi sudah mencapai angka 3000 lebih, lalu untuk kasus di-recovered pada minggu ke-6 mencapai lebih kurang 200 kasus sedangkan kasus kematian mencapai angka lebih dari 200 pada minggu ke-6.

```
In [39]: plt.figure(figsize=(10,6))
plt.plot(week_num_indo,indo_weekwise_confirmed,linewidth=3,label="Kasus di-konfirmasi")
plt.plot(week_num_indo,indo_weekwise_recovered,linewidth=3,label="Kasus di-recovered")
plt.plot(week_num_indo,indo_weekwise_deaths,linewidth=3,label="Kasus meninggal")
plt.xlabel('Week Number')
plt.ylabel("Number of Cases")
plt.title("Perkembangan kasus per minggu")
plt.legend()
```

Out[39]: <matplotlib.legend.Legend at 0x1b23a316f98>



- Keterangan pada Grafik, Kasus perminggu
  - o Grafik ini sama dengan grafik diatas, namun disatukan menjadi satu grafik untuk melihat secara keseluruhan perkembangan kasus COVID-19 di Indonesia.

```
In [40]: fig, (ax1,ax2) = plt.subplots(1, 2, figsize=(15,5))
sns.barplot(x=week_num_indo,y=pd.Series(indo_weekwise_confirmed).diff().fillna(0),color='brown')
sns.barplot(x=week_num_indo,y=pd.Series(indo_weekwise_deaths).diff().fillna(0),color='purple')
ax1.set_xlabel("Week Number")
ax2.set_xlabel("Week Number")
ax1.set_ylabel("Number of Confirmed Cases")
ax2.set_ylabel("Number of Death Cases")
ax1.set_title("Indonesian's Weekwise increase in Number of Confirmed Cases")
ax2.set_title("Indonesians's Weekwise increase in Number of Death Cases")
```

Out[40]: Text(0.5, 1.0, "Indonesians's Weekwise increase in Number of Death Cases")



- Keterangan pada Grafik - Kasus di-konfirmasi dan kasus kematian perminggu
  - o Pada kasus di-konfirmasi grafik menunjukkan peningkatan permingtonya, sedangkan pada grafik kematian, terjadi penurunan di minggu ke-6.

```
In [41]: max_ind=datewise_indo["Confirmed"].max()
plt.figure(figsize=(12,6))
plt.plot(datewise_Italy[(datewise_Ity["Confirmed"]>0)&(datewise_Ity["Confirmed"]<=max]
plt.plot(datewise_US[(datewise_US["Confirmed"]>0)&(datewise_US["Confirmed"]<=max]
plt.plot(datewise_Spain[(datewise_Spain["Confirmed"]>0)&(datewise_Spain["Confirmed"]<=max]
plt.plot(datewise_indo[datewise_indo["Confirmed"]>0]["Confirmed"],label="Confirmed")
plt.xlabel("Date")
plt.ylabel("Number of Confirmed Cases")
plt.title("Growth of Confirmed Cases")
plt.legend()
plt.xticks(rotation=90)

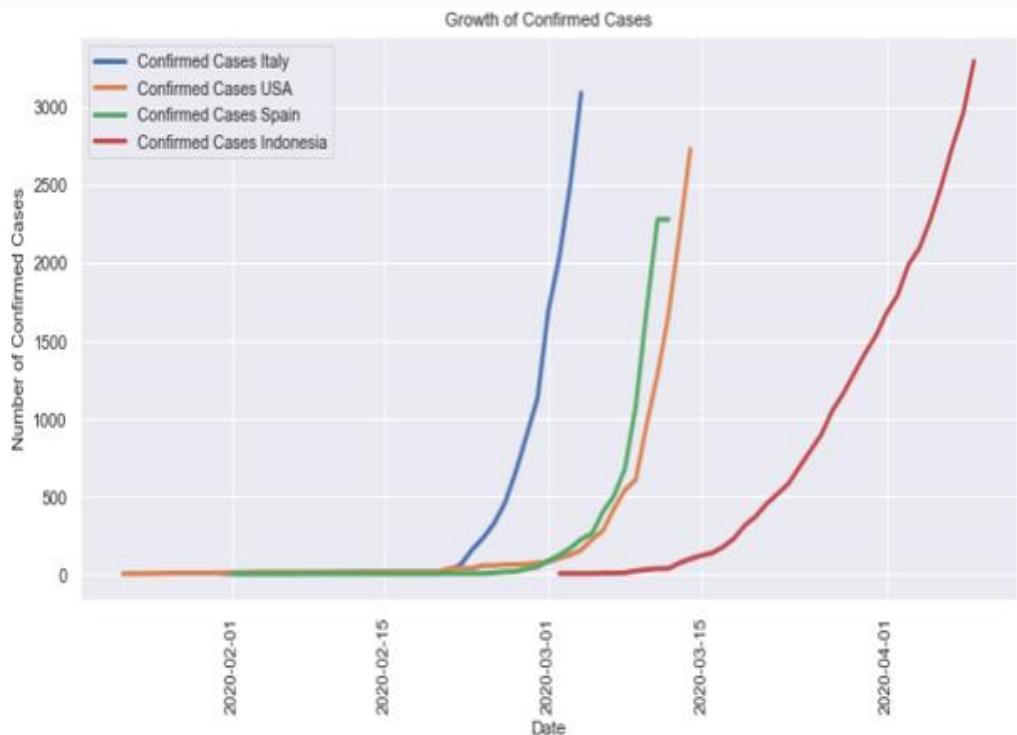
print("It took",datewise_Ity[(datewise_Ity["Confirmed"]>0)&(datewise_Ity["Confirmed"]<=max).shape[0],"days in Italy"]
print("It took",datewise_US[(datewise_US["Confirmed"]>0)&(datewise_US["Confirmed"]<=max).shape[0],"days in USA")
print("It took",datewise_Spain[(datewise_Spain["Confirmed"]>0)&(datewise_Spain["Confirmed"]<=max).shape[0],"days in Spain")
print("It took",datewise_indo[datewise_indo["Confirmed"]>0].shape[0],"days in Indonesia")
```

It took 34 days in Italy to reach number of Confirmed Cases equivalent to Indonesia

It took 53 days in USA to reach number of Confirmed Cases equivalent to Indonesia

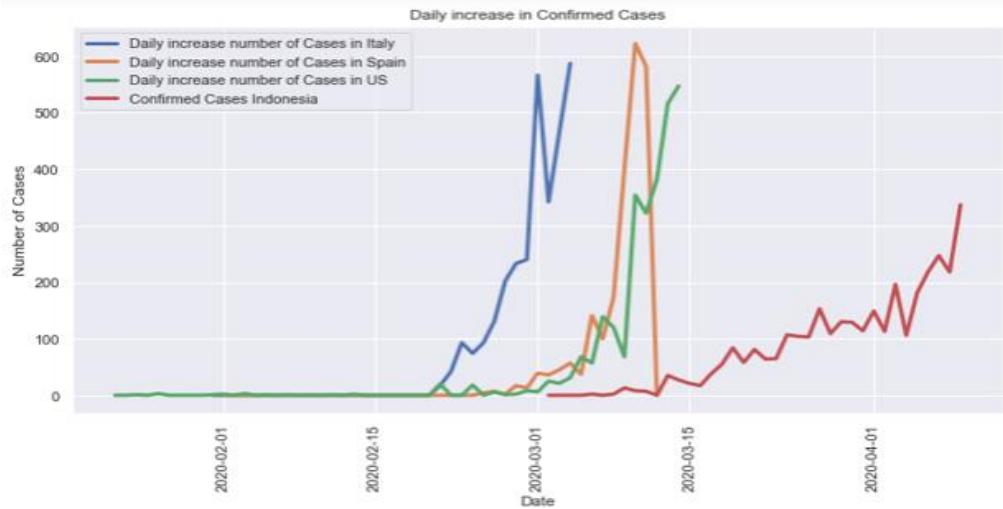
It took 41 days in Spain to reach number of Confirmed Cases equivalent to Indonesia

It took 39 days in Indonesia to reach 3293.0 Confirmed Cases



- Keterangan Grafik - Pertumbuhan kasus di Indonesia
  - o Butuh 39 hari untuk Indonesia mencapai angka 3293 untuk kasus dikonfirmasi, diatas adalah perbandingan dengan Italy, USA dan Spain.

```
In [42]: plt.figure(figsize=(12,6))
plt.plot(datewise_Italy[(datewise_Ity["Confirmed"]>0)&(datewise_Ity["Confirmed"].diff().fillna(0)>0)]
plt.plot(datewise_Spain[(datewise_Spain["Confirmed"]>0)&(datewise_Spain["Confirmed"].diff().fillna(0)>0)])
plt.plot(datewise_US[(datewise_US["Confirmed"]>0)&(datewise_US["Confirmed"]<=max(datewise_US["Confirmed"]))])
plt.plot(datewise_indo[datewise_indo["Confirmed"]>0]["Confirmed"].diff().fillna(0)>0])
plt.ylabel("Number of Cases")
plt.xlabel("Date")
plt.title("Daily increase in Confirmed Cases")
plt.legend()
plt.xticks(rotation=90)
< />
Out[42]: (array([737456., 737470., 737485., 737499., 737516.]),
<a list of 5 Text xticklabel objects>)
```



- Pertumbuhan kasus perhari di Italy, Spain, US dan Indonesia
- Perbandingan pertumbuhan COVID-19 di keempat negara.

```
In [43]: covid_19_INDONESIA = covid[covid['Country/Region'] == 'Indonesia']
covid_19_INDONESIA = covid.groupby("ObservationDate")["Confirmed", "Deaths", "Recovered"]
covid_19_INDONESIA.tail()
< />
Out[43]: ObservationDate Confirmed Deaths Recovered
74 2020-04-05 1272115.0 69374.0 260012.0
75 2020-04-06 1345101.0 74565.0 276515.0
76 2020-04-07 1426096.0 81865.0 300054.0
77 2020-04-08 1511104.0 88338.0 328661.0
78 2020-04-09 1595350.0 95455.0 353975.0
```

- o Angka kasus di Indonesia

- **Analisa Data COVID-19 Indonesia (DATASET INDONESIAN CORONAVIRUS DATASET)**

```
In [44]: cases = pd.read_csv("cases.csv")
confirmed_acc = pd.read_csv("confirmed_acc.csv")
jabar = pd.read_csv("jabar.csv")
jakarta = pd.read_csv("jakarta.csv")
keywordtrend = pd.read_csv("keywordtrend.csv")
patient = pd.read_csv("patient.csv")
province_timeline = pd.read_csv("province_timeline.csv")
```

```
In [45]: cases.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 17 columns):
date 30 non-null object
new_tested 28 non-null float64
acc_tested 30 non-null int64
new_confirmed 30 non-null int64
acc_confirmed 30 non-null int64
acc_negative 30 non-null int64
being_checked 30 non-null int64
isolated 30 non-null int64
new_released 30 non-null int64
acc_released 30 non-null int64
new_deceased 30 non-null int64
acc_deceased 30 non-null int64
positive_rate 30 non-null object
negative_rate 30 non-null object
decease_rate 30 non-null object
release_rate 30 non-null object
dailypositive_rate 30 non-null object
dtypes: float64(1), int64(10), object(6)
memory usage: 4.1+ KB
```

○ Ketereangan kolom dataset Indonesia

|    |                   |       |      |    |     |      |    |     |   |   |   |
|----|-------------------|-------|------|----|-----|------|----|-----|---|---|---|
| 7  | 9-<br>Mar-<br>20  | 60.0  | 543  | 13 | 19  | 487  | 37 | 19  | 0 | 0 | 0 |
| 8  | 10-<br>Mar-<br>20 | 151.0 | 694  | 8  | 27  | 648  | 19 | 27  | 0 | 0 | 0 |
| 9  | 11-<br>Mar-<br>20 | 99.0  | 793  | 7  | 34  | 744  | 15 | 30  | 2 | 2 | 2 |
| 10 | 12-<br>Mar-<br>20 | 69.0  | 862  | 0  | 34  | 811  | 17 | 27  | 1 | 3 | 2 |
| 11 | 13-<br>Mar-<br>20 | 143.0 | 1005 | 35 | 69  | 917  | 19 | 60  | 2 | 5 | 0 |
| 12 | 14-<br>Mar-<br>20 | 200.0 | 1205 | 27 | 96  | 1109 | 0  | 83  | 3 | 8 | 1 |
| 13 | 15-<br>Mar-<br>20 | 88.0  | 1293 | 21 | 117 | 1167 | 9  | 104 | 0 | 8 | 0 |
| 14 | 16-<br>Mar-<br>20 | NaN   | 1330 | 17 | 134 | 1183 | 13 | 121 | 0 | 8 | 0 |

| In [46]: | cases     |            |            |               |               |              |               |          |              |              |              |          |
|----------|-----------|------------|------------|---------------|---------------|--------------|---------------|----------|--------------|--------------|--------------|----------|
| Out[46]: | date      | new_tested | acc_tested | new_confirmed | acc_confirmed | acc_negative | being_checked | isolated | new_released | acc_released | new_deceased | acc_dece |
| 0        | 2-Mar-20  | NaN        | 339        | 2             | 2             | 335          | 2             | 2        | 0            | 0            | 0            | 0        |
| 1        | 3-Mar-20  | 2.0        | 341        | 0             | 2             | 337          | 2             | 2        | 0            | 0            | 0            | 0        |
| 2        | 4-Mar-20  | 31.0       | 372        | 0             | 2             | 356          | 14            | 2        | 0            | 0            | 0            | 0        |
| 3        | 5-Mar-20  | 16.0       | 388        | 0             | 2             | 371          | 15            | 2        | 0            | 0            | 0            | 0        |
| 4        | 6-Mar-20  | 62.0       | 450        | 2             | 4             | 422          | 24            | 4        | 0            | 0            | 0            | 0        |
| 5        | 7-Mar-20  | 4.0        | 454        | 0             | 4             | 422          | 28            | 4        | 0            | 0            | 0            | 0        |
| 6        | 8-Mar-20  | 29.0       | 483        | 2             | 6             | 445          | 32            | 6        | 0            | 0            | 0            | 0        |
| 15       | 17-Mar-20 | 142.0      | 1372       | 38            | 172           | 1278         | 22            | 156      | 1            | 9            | 2            |          |
| 16       | 18-Mar-20 | 220.0      | 1592       | 55            | 227           | 1342         | 23            | 197      | 2            | 11           | 12           |          |
| 17       | 19-Mar-20 | 306.0      | 1651       | 82            | 309           | 1570         | 19            | 269      | 4            | 15           | 6            |          |
| 18       | 20-Mar-20 | 130.0      | 2028       | 60            | 369           | 1643         | 18            | 321      | 1            | 16           | 7            |          |
| 19       | 21-Mar-20 | 337.0      | 2365       | 81            | 450           | 1895         | 20            | 392      | 4            | 20           | 6            |          |
| 20       | 22-Mar-20 | 73.0       | 2438       | 64            | 514           | 1904         | 20            | 437      | 9            | 29           | 10           |          |
| 21       | 23-Mar-20 | 318.0      | 2756       | 65            | 579           | 2177         | 0             | 500      | 1            | 30           | 1            |          |
| 22       | 24-Mar-20 | 576.0      | 3332       | 106           | 685           | 2625         | 22            | 600      | 0            | 30           | 6            |          |
| 23       | 25-Mar-20 | 490.0      | 3822       | 105           | 790           | 3032         | 0             | 701      | 1            | 31           | 3            |          |
| 24       | 26-Mar-20 | 514.0      | 4336       | 103           | 893           | 3443         | 0             | 780      | 4            | 35           | 20           |          |
| 25       | 27-Mar-20 | 1439.0     | 5775       | 153           | 1046          | 4729         | 0             | 913      | 11           | 46           | 9            |          |
| 26       | 28-Mar-20 | 491.0      | 6266       | 109           | 1155          | 5111         | 0             | 994      | 13           | 59           | 15           |          |
| 27       | 29-Mar-20 | 268.0      | 6534       | 130           | 1285          | 5249         | 0             | 1107     | 5            | 64           | 12           |          |
| 28       | 30-Mar-20 | 129.0      | 6663       | 129           | 1414          | 5249         | 0             | 1217     | 11           | 75           | 8            |          |
| 29       | 31-Mar-20 | 114.0      | 6777       | 114           | 1528          | 5249         | 0             | 1311     | 6            | 81           | 14           |          |

## ○ Dataset Indonesia

```
In [47]: fig = plt.figure(figsize=(16,8))
ax = fig.add_subplot(111)
cases.groupby('date').mean().sort_values(by='acc_tested', ascending=False)['acc_tested'].plot('bar', color='r',width=0.3,title='Test dilakukan perhari')
plt.xticks(rotation = 90)
plt.ylabel('acc_tested')
ax.title.set_fontsize(30)
ax.xaxis.label.set_fontsize(10)
ax.yaxis.label.set_fontsize(10)
print(cases.groupby('date').mean().sort_values(by='acc_tested', ascending=False)[['acc_tested']][[1,2]])
print(cases.groupby('date').mean().sort_values(by='acc_tested', ascending=False)[['acc_tested']][[4,5,6]])

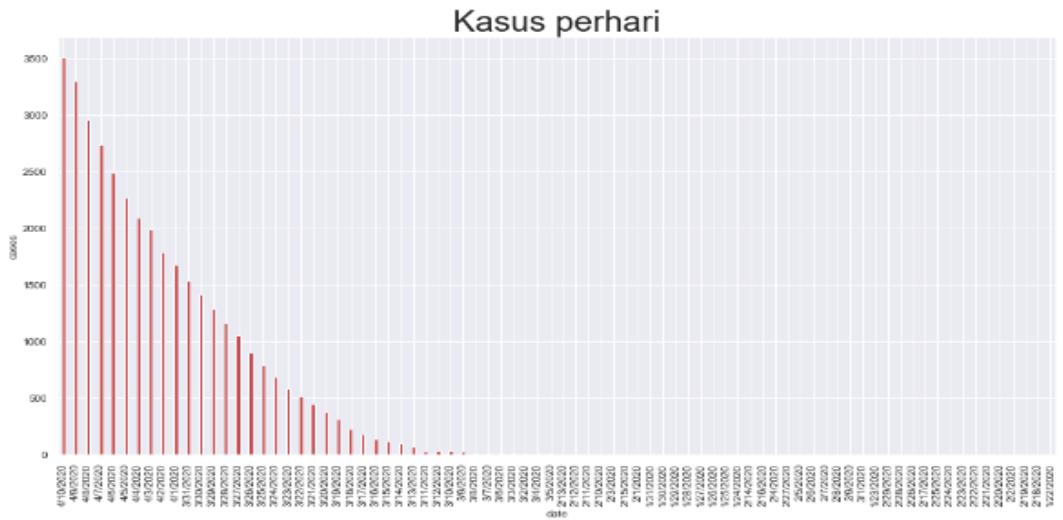
```

date  
 30-Mar-20 6663  
 29-Mar-20 6534  
 Name: acc\_tested, dtype: int64  
 date  
 27-Mar-20 5775  
 26-Mar-20 4336  
 25-Mar-20 3822  
 Name: acc\_tested, dtype: int64



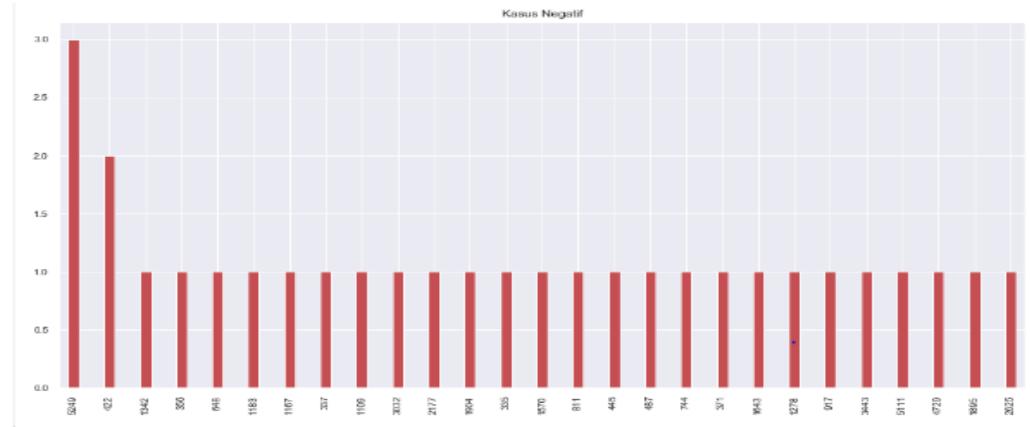
- Keterangan Grafik - Daily Test

- Drafik pengadaan test perhari di indonesia pertanggal 31 Maret 2020.
- Data PDP di Jakarta pada bulan maret 2020, dapat dilihat dari grafik diatas, penaikan terus terjadi sampai tanggal 18 maret 2020 penaikan terhenti satu hari, dan besoknya penaikan total PDP di Jakarta melonjak jauh, dan pada tanggal 24, 25, 26 maret 2020 total PDP mencapai lebih 800.



- Keterangan Grafik - Kasus perhari
  - o Kasus perhari sampai tanggal 9 Maret 2020 di indonesia.

```
In [49]: fig = plt.figure(figsize=(16,8))
cases.acc_negative.value_counts().plot(kind='bar',color='r',width=0.3,title='Kasus Negatif')
plt.show()
```



- Keterangan Grafik - Kasus Negatif
  - o Kasus negatif di indonesia.

In [58]: jakarta

Out[58]:

|           | date  | odp_process | odp_end | odp_total | pdp_process | pdp_end | pdp_total | jakbar | jakpus | jaksel | jaktim | jakut | outsidejakarta | unknown | confirmed | releas |
|-----------|-------|-------------|---------|-----------|-------------|---------|-----------|--------|--------|--------|--------|-------|----------------|---------|-----------|--------|
| 3/1/2020  | 14.0  | 115.0       | 129.0   | 9.0       | 30.0        | 39.0    | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/2/2020  | 14.0  | 115.0       | 129.0   | 9.0       | 30.0        | 39.0    | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/3/2020  | 94.0  | 121.0       | 215.0   | 24.0      | 30.0        | 54.0    | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/4/2020  | 120.0 | 121.0       | 241.0   | 26.0      | 30.0        | 56.0    | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/5/2020  | 127.0 | 140.0       | 267.0   | 49.0      | 34.0        | 83.0    | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/6/2020  | 83.0  | 221.0       | 304.0   | 44.0      | 44.0        | 88.0    | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/7/2020  | 94.0  | 262.0       | 356.0   | 44.0      | 54.0        | 98.0    | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/8/2020  | 64.0  | 298.0       | 362.0   | 57.0      | 70.0        | 127.0   | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/9/2020  | 68.0  | 310.0       | 378.0   | 87.0      | 79.0        | 166.0   | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/10/2020 | 70.0  | 331.0       | 401.0   | 97.0      | 100.0       | 197.0   | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/11/2020 | 98.0  | 347.0       | 445.0   | 103.0     | 116.0       | 219.0   | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/12/2020 | 112.0 | 374.0       | 486.0   | 120.0     | 118.0       | 238.0   | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/13/2020 | 144.0 | 442.0       | 586.0   | 126.0     | 135.0       | 261.0   | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |
| 3/14/2020 | 174.0 | 486.0       | 660.0   | 152.0     | 137.0       | 289.0   | NaN       | NaN    | NaN    | NaN    | NaN    | NaN   | NaN            | NaN     | NaN       | NaN    |

```
In [48]: fig = plt.figure(figsize=(16,8))
ax = fig.add_subplot(111)
confirmed_acc.groupby('date').mean().sort_values(by='cases', ascending=False)['cases'].plot('bar', color='r', width=0.3, title='K
plt.xticks(rotation = 90)
plt.ylabel('cases')
ax.title.set_fontsize(30)
ax.xaxis.label.set_fontsize(10)
ax.yaxis.label.set_fontsize(10)
print(confirmed_acc.groupby('date').mean().sort_values(by='cases', ascending=False)[['cases']][[1,2]])
print(confirmed_acc.groupby('date').mean().sort_values(by='cases', ascending=False)[['cases']][[4,5,6]])
< >
```

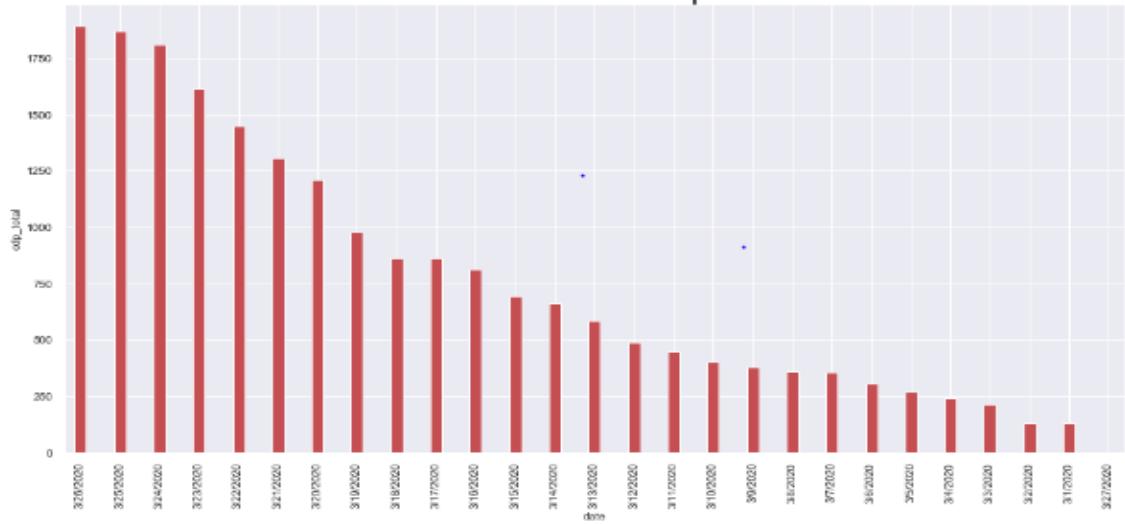
|                           |          |      |
|---------------------------|----------|------|
| date                      | 4/9/2020 | 3293 |
|                           | 4/8/2020 | 2956 |
| Name: cases, dtype: int64 |          |      |
| date                      | 4/6/2020 | 2491 |
|                           | 4/5/2020 | 2273 |
|                           | 4/4/2020 | 2092 |
| Name: cases, dtype: int64 |          |      |

## ○ Data Covid-19 untuk Jakarta

```
In [51]: fig = plt.figure(figsize=(16,8))
ax = fig.add_subplot(111)
jakarta.groupby('date').mean().sort_values(by='odp_total', ascending=False)['odp_total'].plot('bar', color='r', width=0.3, title='
plt.xticks(rotation = 90)
plt.ylabel('odp_total')
ax.title.set_fontsize(30)
ax.xaxis.label.set_fontsize(10)
ax.yaxis.label.set_fontsize(10)
print(jakarta.groupby('date').mean().sort_values(by='odp_total', ascending=False)[['odp_total']][[1,2]])
print(jakarta.groupby('date').mean().sort_values(by='odp_total', ascending=False)[['odp_total']][[4,5,6]])
< >
```

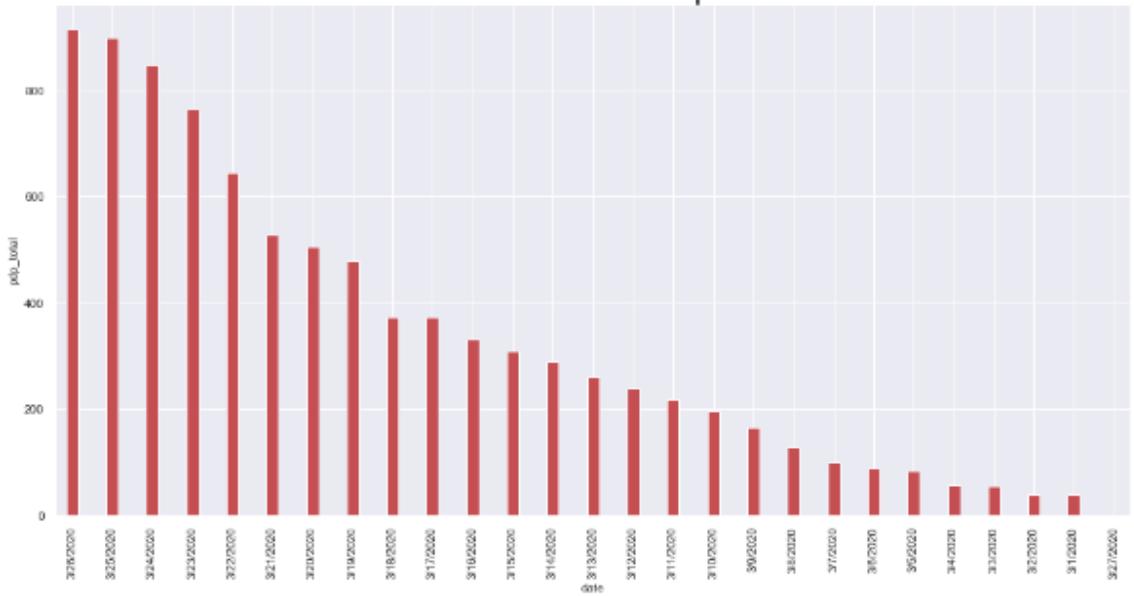
|                                 |           |        |
|---------------------------------|-----------|--------|
| date                            | 3/25/2020 | 1872.0 |
|                                 | 3/24/2020 | 1811.0 |
| Name: odp_total, dtype: float64 |           |        |
| date                            | 3/22/2020 | 1447.0 |
|                                 | 3/21/2020 | 1306.0 |
|                                 | 3/20/2020 | 1209.0 |
| Name: odp_total, dtype: float64 |           |        |

### Total ODP Jakarta perhari



- Keterangan Grafik - ODP Jakarta
  - o Grafik pertumbuhan ODP di Jakarta.

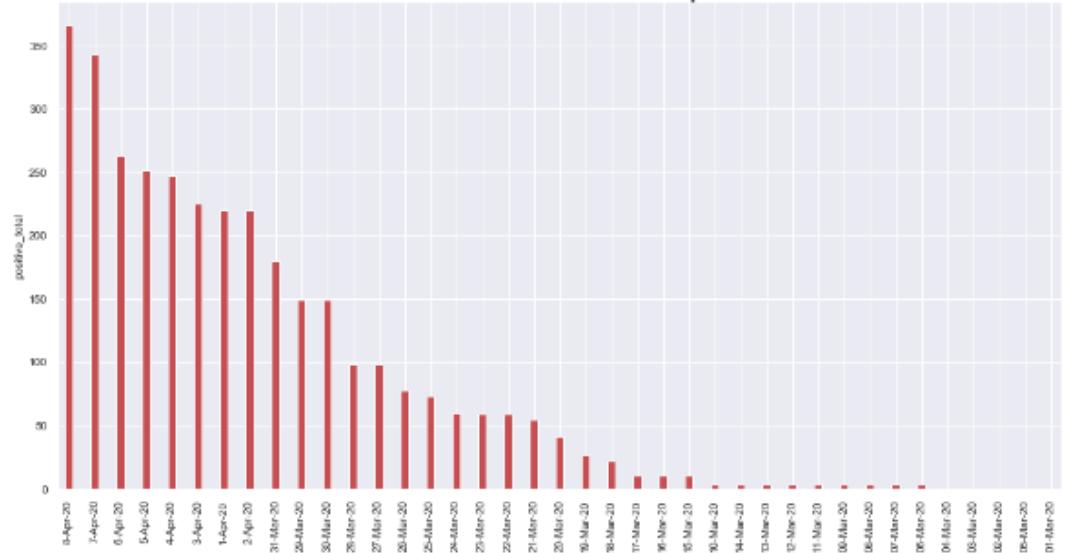
### Total PDP Jakarta perhari



- Keterangan Grafik - PDP jakarta
  - o Data PDP di Jakarta diambil pada bulan Maret 2020.

- Data Covid-19 Untuk Jawa Barat

Total Positif Jawa Barat perhari



- Keterangan Grafik - Kasus positif di Jawa Barat

- o Perkembangan kasus positif di Jawa Barat perhari yang terus meningkat
- o Perkembangan kasus postif di Jawa Barat Perhari, dapat dilihat dari grafik diatas, penaikan yang melonjak tinggi dimulai pada tanggal 30 maret 2020 yang mencapai hampir 150 pasien yang positif. Dan pada tanggal 7 sampai 8 april, penaikan melonjak jauh mencapai 300 lebih pasien.

```
In [55]: fig = plt.figure(figsize=(16,8))
ax = fig.add_subplot(111)
jabar.groupby('date').mean().sort_values(by='odp_total', ascending=False)['odp_total'].plot('bar', color='r',width=0.3,title='Total ODP Jawa Barat perhari')
plt.xticks(rotation = 90)
plt.ylabel('odp_total')
ax.title.set_fontsize(30)
ax.xaxis.label.set_fontsize(10)
ax.yaxis.label.set_fontsize(10)
print(jabar.groupby('date').mean().sort_values(by='odp_total', ascending=False)[['odp_total'][[1,2]]])
print(jabar.groupby('date').mean().sort_values(by='odp_total', ascending=False)[['odp_total'][[4,5,6]]])
<ipython-input-55-133a2a2a2a2>
```

| Date                          | odp_total |
|-------------------------------|-----------|
| 8-Apr-20                      | 21991     |
| 7-Apr-20                      | 21991     |
| 6-Apr-20                      | 21991     |
| Name: odp_total, dtype: int64 |           |
| date                          |           |
| 4-Apr-20                      | 20712     |
| 3-Apr-20                      | 16932     |
| 1-Apr-20                      | 14291     |
| Name: odp_total, dtype: int64 |           |



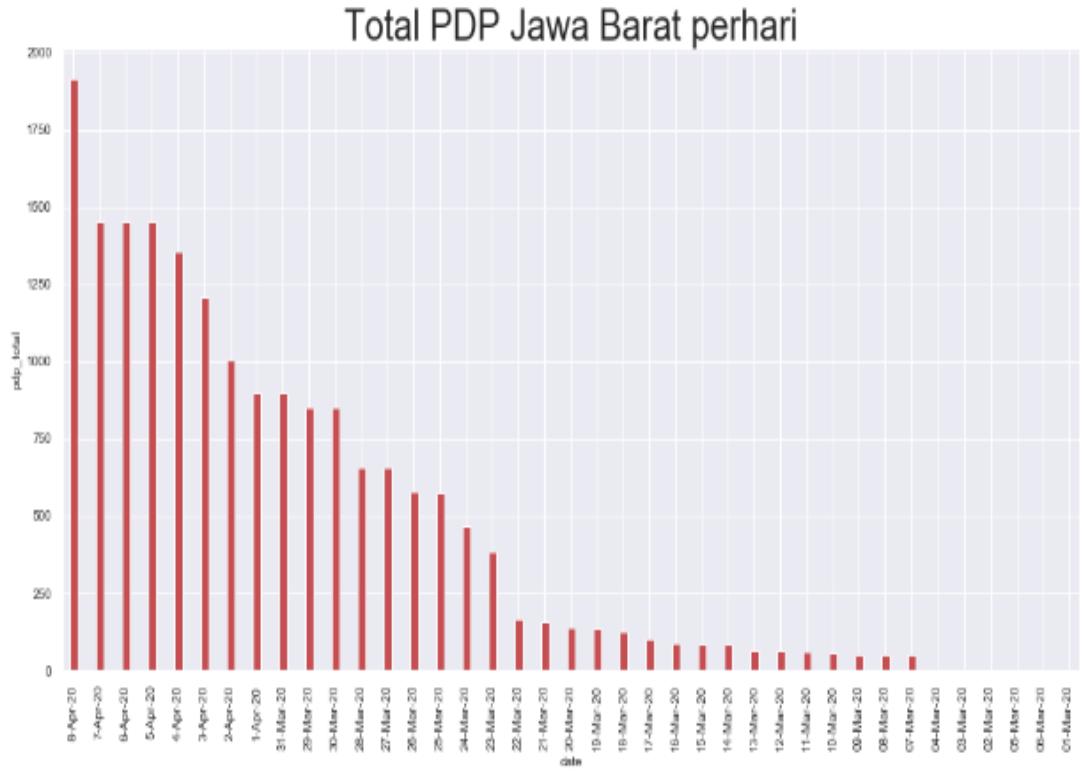
#### - Keterangan Grafik - Total ODP Jawa Barat

- Total ODP di Jawa Barat, dilihat dari grafik makin hari ODP di provinsi jawa barat makin meningkat.
- Perkembangan total ODP di Jawa Barat dari tanggal 1 maret sampai 8 april 2020, bisa dilihat pelonjakan pertama dimulai pada tanggal 29-30 maret 2020 sebanyak 5000 ODP. Dan besoknya pada tanggal 31 maret 2020, jumlah ODP terjadi kenaikan menjadi 10000 ODP. Dan pada tanggal 4,5,6, dan 7 april terjadi kenaikan menjadi 20000 ODP, dan besoknya tanggal 8 april 2020, pasien ODP melonjak tinggi menjadi 25000 ODP.

```
In [56]: fig = plt.figure(figsize=(16,8))
ax = fig.add_subplot(111)
jabar.groupby('date').mean().sort_values(by='pdp_total', ascending=False)[['pdp_total']].plot('bar', color='r',width=0.3,title='Total PDP Jawa Barat perhari')
plt.xticks(rotation = 90)
plt.ylabel('pdp_total')
ax.title.set_fontsize(30)
ax.xaxis.label.set_fontsize(10)
ax.yaxis.label.set_fontsize(10)
print(jabar.groupby('date').mean().sort_values(by='pdp_total', ascending=False)[['pdp_total']][[1,2]])
print(jabar.groupby('date').mean().sort_values(by='pdp_total', ascending=False)[['pdp_total']][[4,5,6]])
<ipython-input-56-133a2a2a2a>
```

date

| Date                          | pdp_total |
|-------------------------------|-----------|
| 8-Apr-20                      | 1451      |
| 7-Apr-20                      | 1451      |
| 6-Apr-20                      | 1451      |
| Name: pdp_total, dtype: int64 |           |
| date                          |           |
| 4-Apr-20                      | 1354      |
| 3-Apr-20                      | 1205      |
| 2-Apr-20                      | 1003      |
| Name: pdp_total, dtype: int64 |           |



- Keterangan Grafik - Total PDP Jawa Barat

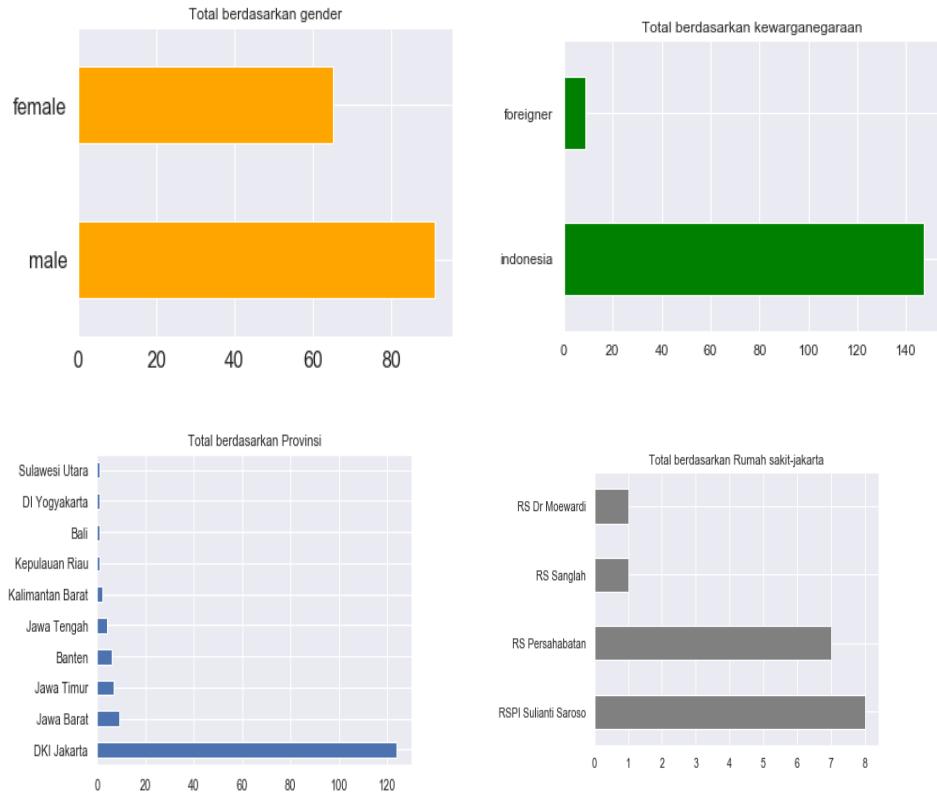
- o Total PDP di Jawa Barat, dilihat dari grafik makin hari PDP di provinsi jawa barat makin meningkat.
- o Perkembang total PDP di Jawa Barat dari 1 maret sampai 8 april 2020, bisa dilihat pada grafik diatas, total PDP melonjok pada awak tanggal 23 maret 2020 dan terus menaik sampai tanggal 27 maret 2020, sedangkan pada tanggal 28 maret total PDP berjumlah lebih 500 PDP, besoknya pada tanggal 29 - 30 maret 2020, jumlah PDP naik lebih dari

750 PDP. Sedangkan pada tanggal 1-2 april 2020, pelonjakan PDP tidak terlalu tinggi. Lalu pada tanggal 3 april 2020, jumlah PDP menyentuh angka 1000 dan terus naik hingga pada 7 april 2020 jumlah PDP terdapat lebih dari 1250 PDP. Dan paling mengejutkan pada tanggal 8 april 2020, pelonjakan jumlah PDP meningkat tinggi, yaitu berjumlah lebih dari 1750 PDP.

| In [57]: patient |        |        |             |           |               |                |                |               |               |          |                    |
|------------------|--------|--------|-------------|-----------|---------------|----------------|----------------|---------------|---------------|----------|--------------------|
| patient_id       | gender | age    | nationality | province  | current_state | contacted_with | confirmed_date | released_date | deceased_date | hospital |                    |
| 0                | 1      | female | 31.0        | indonesia | DKI Jakarta   | released       | NaN            | 2-Mar-20      | 13-Mar-20     | NaN      | RSPI Sultan Saroso |
| 1                | 2      | female | 64.0        | indonesia | DKI Jakarta   | released       | 1.0            | 2-Mar-20      | 16-Mar-20     | NaN      | RSPI Sultan Saroso |
| 2                | 3      | female | 33.0        | indonesia | DKI Jakarta   | released       | 1.0            | 6-Mar-20      | 13-Mar-20     | NaN      | RSPI Sultan Saroso |
| 3                | 4      | female | 34.0        | indonesia | DKI Jakarta   | isolated       | 1.0            | 6-Mar-20      | NaN           | NaN      | RSPI Sultan Saroso |
| 4                | 5      | male   | 55.0        | indonesia | DKI Jakarta   | isolated       | 1.0            | 8-Mar-20      | NaN           | NaN      | RS Persahabatan    |
| ...              | ...    | ...    | ...         | ...       | ...           | ...            | ...            | ...           | ...           | ...      | ...                |
| 888              | 889    | NaN    | NaN         | NaN       | NaN           | NaN            | NaN            | 26-Mar-20     | NaN           | NaN      | NaN                |
| 889              | 890    | NaN    | NaN         | NaN       | NaN           | NaN            | NaN            | 26-Mar-20     | NaN           | NaN      | NaN                |
| 890              | 891    | NaN    | NaN         | NaN       | NaN           | NaN            | NaN            | 26-Mar-20     | NaN           | NaN      | NaN                |
| 891              | 892    | NaN    | NaN         | NaN       | NaN           | NaN            | NaN            | 26-Mar-20     | NaN           | NaN      | NaN                |
| 892              | 893    | NaN    | NaN         | NaN       | NaN           | NaN            | NaN            | 26-Mar-20     | NaN           | NaN      | NaN                |

893 rows × 11 columns

```
In [58]: patient.gender.value_counts().plot(kind='barh',color='orange',title='Total berdasarkan gender',fontsize=16)
plt.show()
patient.nationality.value_counts().plot(kind='barh',color='green', title='Total berdasarkan kewarganegaraan')
plt.show()
patient.province.value_counts().plot(kind='barh',title='Total berdasarkan Provinsi')
plt.show()
patient.hospital.value_counts().plot(kind='barh',color='grey', title='Total berdasarkan Rumah sakit-jakarta')
plt.show()
```



- Keterangan Grafik - Sebaran data
  - o Sebaran data gender, kewarganegaraan, provinsi dan rumah sakit kasus COVID-19 di Indonesia.
- Total Bedasarkan Gender
  - o Dapat dilihat grafik diagram diatas, menunjukkan bahwa diagram pasien yang terkonfirmasi paling banyak adalah laki laki dibandingkan perempuan. Laki laki mencapai lebih dari 80, sedangkan perempuan lebih dari 60 dan kurang dari 80.
- Total Bedasarkan Kewarganegaraan
  - o Bedasarkan grafik diagram diatas, menunjukkan jumlah pasien yang terkonfirmasi yang berasal dari negara Indonesia lebih banyak dibandingkan pasien yang berasal dari negara asing. Pasien yang berasal dari Indonesia melebihi 140. Sedangkan untuk pasien dari negara asing kurang dari 20.
- Total Bedasarkan Provinsi
  - o Bedasarkan grafik diagram diatas, menunjukkan jumlah pasien dari berbagai provinsi. Provinsi yang paling banyak pasien yang terkonfirmasi adalah provinsi DKI Jakarta dengan jumlah lebih 120, sedangkan untuk empat provinsi yaitu Sulawesi Utara, DI Yogyakarta, Bali, Kepulaun Riau, jumlah pasiennya sedikit, kurang dari 20.
- Total Bedasarkan Rumah Sakit-Jakarta
  - o Bedasarkan grafik diagram diatas, menunjukkan jumlah total bedasarkan rumah sakit-Jakarta. Yang paling banyak adalah rumah sakit RSPI Sulianti Saroso dengan jumlah mencapai 8.

## 6. MACHINE LEARNING

### A. LINIER REGRESSION

```
1 datewise["Days Since"] = datewise.index - datewise.index[0]
2 datewise["Days Since"] = datewise["Days Since"].dt.days

1 train_ml = datewise.iloc[:int(datewise.shape[0]*0.90)]
2 valid_ml = datewise.iloc[int(datewise.shape[0]*0.90):]
3 model_scores = []
```

- Membuat data train dan validasi untuk Linier Regression. Train diambil dari tanggal 22 Januari sampai 01 April. Sedangkan untuk validasi diambil tanggal 02 April sampai 09 April. Dataset sudah di preprocessing sebelumnya untuk memilih kolom” yang akan dimasukan kedalam Machine Learning.

```
1 train_ml
```

|                 | Confirmed | Recovered | Deaths  | Days Since |
|-----------------|-----------|-----------|---------|------------|
| ObservationDate |           |           |         |            |
| 2020-01-22      | 555.0     | 28.0      | 17.0    | 0          |
| 2020-01-23      | 653.0     | 30.0      | 18.0    | 1          |
| 2020-01-24      | 941.0     | 36.0      | 26.0    | 2          |
| 2020-01-25      | 1438.0    | 39.0      | 42.0    | 3          |
| 2020-01-26      | 2118.0    | 52.0      | 56.0    | 4          |
| ...             | ...       | ...       | ...     | ...        |
| 2020-03-28      | 660693.0  | 139415.0  | 30852.0 | 66         |
| 2020-03-29      | 720140.0  | 149082.0  | 33925.0 | 67         |
| 2020-03-30      | 782389.0  | 164566.0  | 37582.0 | 68         |
| 2020-03-31      | 857487.0  | 178034.0  | 42107.0 | 69         |
| 2020-04-01      | 932605.0  | 193177.0  | 46809.0 | 70         |

71 rows × 4 columns

```
1 valid_ml
```

|                 | Confirmed | Recovered | Deaths  | Days Since |
|-----------------|-----------|-----------|---------|------------|
| ObservationDate |           |           |         |            |
| 2020-04-02      | 1013303.0 | 210263.0  | 52983.0 | 71         |
| 2020-04-03      | 1095917.0 | 225798.0  | 58787.0 | 72         |
| 2020-04-04      | 1197408.0 | 246152.0  | 64806.0 | 73         |
| 2020-04-05      | 1272115.0 | 260012.0  | 69374.0 | 74         |
| 2020-04-06      | 1345101.0 | 276515.0  | 74865.0 | 75         |
| 2020-04-07      | 1426096.0 | 300054.0  | 81865.0 | 76         |
| 2020-04-08      | 1511104.0 | 328661.0  | 88338.0 | 77         |
| 2020-04-09      | 1595350.0 | 353975.0  | 95455.0 | 78         |

- Data train. & Validasi untuk proses training.

```

1 lin_reg=LinearRegression(normalize=True)

1 lin_reg.fit(np.array(train_ml["Days Since"]).reshape(-1,1),np.array(train_ml["Confirmed"]).reshape(-1,1))
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=True)

1 prediction_valid_linreg=lin_reg.predict(np.array(valid_ml["Days Since"]).reshape(-1,1))
2 prediction_valid_linreg

array([[457547.13400402],
 [465742.75744467],
 [473938.38088531],
 [482134.00432596],
 [490329.6277666],
 [498525.25120724],
 [506720.87464789],
 [514916.49808853]])

```

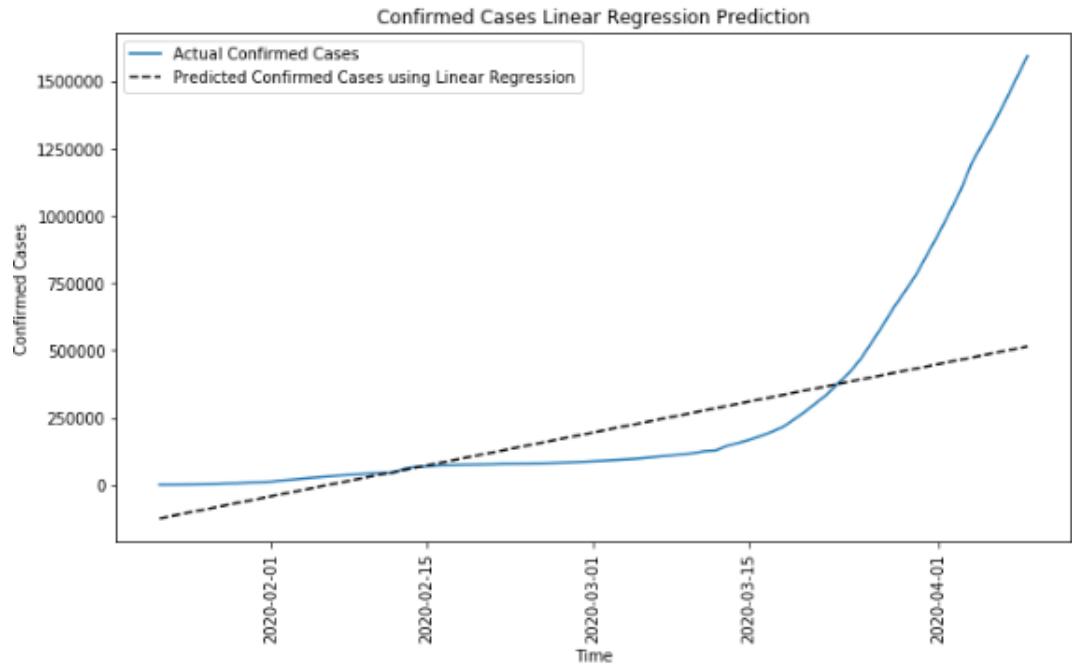
- Normalisasi data train dan validasi dengan kolom Day since untuk di prediksi.

```

1 model_scores.append(np.sqrt(mean_squared_error(valid_ml["Confirmed"],prediction_valid_linreg)))
2 print("Root Mean Square Error for Linear Regression: ",np.sqrt(mean_squared_error(valid_ml["Confirmed"],prediction_
< >
Root Mean Square Error for Linear Regression: 838197.9203347286

```

- Hasil RMSE dari Linier Regression **838197.9203347286**, Angka ini termasuk angka yang buruk untuk prediksi.



- Prediksi Linear Regression Model untuk kasus Confirmed terlihat tidak sama sekali linier atau sesuai dengan data aktual.

## B. SUPPORT VECTOR MACHINE

```
1 #Inisialisasi SVR Model dengan hyperparameters untuk GridSearchCV
2 svm=SVR(C=1,degree=6,kernel='poly',epsilon=0.01)

1 svm
SVR(C=1, cache_size=200, coef0=0.0, degree=6, epsilon=0.01,
gamma='auto_deprecated', kernel='poly', max_iter=-1, shrinking=True,
tol=0.001, verbose=False)
```

### - Keterangan Learning SVM

- o **C : float, optional (default=1.0)**

C merupakan parameter regularisasi, bisa diisi dengan float atau pilihan lain tetapi 1.0 adalah nilai bawaan. Kekuatan regularisasi berbanding terbalik dengan C. Harus benar-benar positif.

- o **Cache\_size : float, optional**

Cache\_size merupakan parameter pengaturan untuk mengatur ukuran cache (dalam MB), bisa bertipe float atau opsi lainnya

- o **Coef0 : float, optional (default = 0.0)**

Istilah independen dalam fungsi kernel, coef0 bisa bertipe float maupun opsi yang lain tetapi memiliki nilai default = 0.0. Ini hanya akan signifikan dalam ‘poli’ dan ‘sigmoid’.

- o **Degree : int, optional (default = 3)**

Degree merupakan tingkat fungsi kernel polinomial ('poli'), bisa bertipe integer atau opsi lainnya tetapi memiliki nilai default = 3. Diabaikan oleh semua kernel lainnya.

- o **Epsilon : float, optional (default = 0.1)**

Epsilon dalam model epsilon-SVR. Ini menentukan tabung epsilon di mana tidak ada penalti yang terkait dalam fungsi kehilangan pelatihan dengan poin yang diprediksi dalam jarak epsilon dari nilai aktual. Epsilon bisa bertipe float atau pilihan lainnya tetapi memiliki nilai default = 0.1

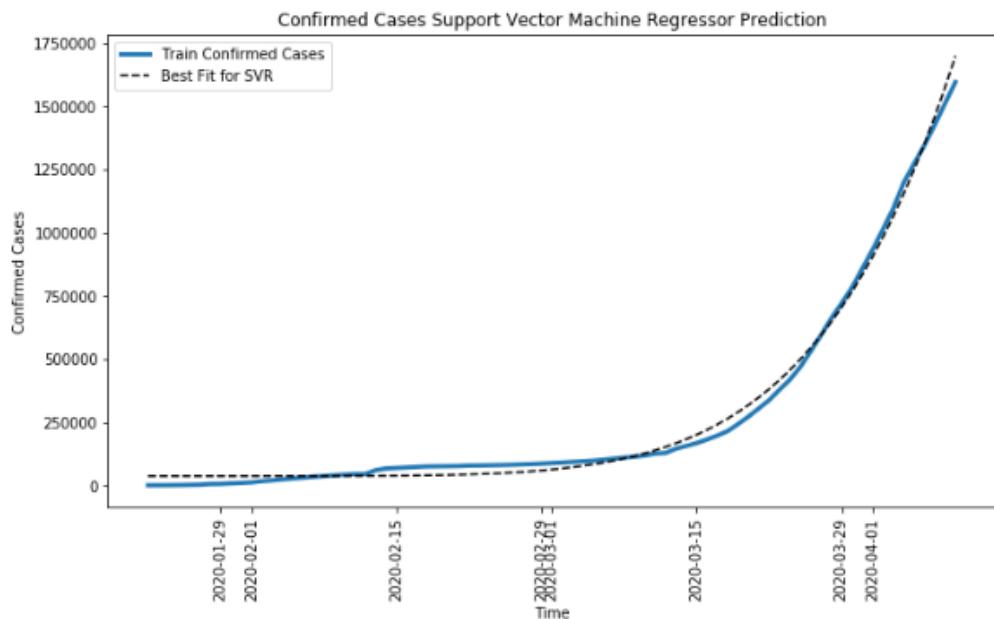
- **Gamma : {'scale','auto'}, float, optional (default 'scale')**  
Merupakan Koefisien kernel untuk 'rbf', 'poly' dan 'sigmoid'. Jika gama = scale (default) maka  $1 / (\text{n\_features} * \text{X.var}())$  sebagai nilai gamma, tetapi jika gamma = auto maka akan menggunakan  $1 / \text{n\_fitur}$ .
- **Kernel : string, optional (default 'rbf')**  
Menentukan jenis kernel yang akan digunakan dalam algoritma. Itu harus salah satu dari 'linear', 'poly', 'rbf', 'sigmoid', 'precomputed' atau callable. Jika tidak ada yang diberikan, 'rbf' akan digunakan. Jika callable diberikan, ini digunakan untuk melakukan precompute terhadap matriks kernel.
- **Max\_iter : int, optional (default = -1)**  
Merupakan Batas keras pada iterasi dalam solver, atau -1 tanpa batas. Memiliki tipe integer ataupun opsi lainnya tetapi nilai defaultnya = -1
- **Shrinking : Boolean, optional (default = true)**  
Shrinking merupakan pilihan Apakah akan menggunakan heuristic menyusut. Opsi default dari shrinking adalah True tetapi masih ada opsi lainnya seperti Boolean dll.
- **Tol : float, optional (default = 1e-3)**  
Tol merupakan Toleransi untuk menghentikan kriteria.
- **Verbose : bool, default = false**  
Aktifkan keluaran verbose. Perhatikan bahwa pengaturan ini mengambil keuntungan dari pengaturan runtime per-proses di libsvm yang, jika diaktifkan, mungkin tidak berfungsi dengan baik dalam konteks multithreaded.

```
1 #Melakukan GridSearchCV untuk menemukan Estimator terbaik
2 svm.fit(np.array(train_ml["Days Since"]).reshape(-1,1),np.array(train_ml["Confirmed"]).reshape(-1,1))
```

- GridsearchCV untuk menemukan Estimator terbaik untuk Machine Learning Support Vector Machine.

```
1 prediction_valid_svm=svm.predict(np.array(valid_ml["Days Since"]).reshape(-1,1))
2 model_scores.append(np.sqrt(mean_squared_error(valid_ml["Confirmed"],prediction_valid_svm)))
3 print("Root Mean Square Error for Support Vectore Machine: ",np.sqrt(mean_squared_error(valid_ml["Confirmed"],predi
< >
Root Mean Square Error for Support Vectore Machine: 50167.10314784511
```

- Hasil RMSE dari SVM **50167.10314784511**, SVM jauh lebih baik daripada Linier Regression.



- Prediksi SVM mendekati data actual untuk kasus confirmed dan akurasi SVM jauh lebih tinggi dibanding Linier Regression.

### C. PERBANDINGAN LINIER REGRESSION DAN SVM

```
1 new_date=[]
2 for i in range(1,18):
3 new_date.append(datewise.index[-1]+timedelta(days=i))
```

- Membuat tanggal prediksi

```
1 pd.set_option('display.float_format', lambda x: '%.6f' % x)
2 model_predictions=pd.DataFrame(zip(new_date,new_prediction_lr,new_prediction_svm),columns=["Dates","Linear Regression Prediction","SVM Prediction"])
3 model_predictions.head()
```

- Membuat table perbandingan Linier Regression dan SVM

|   | Dates      | Linear Regression Prediction | SVM Prediction |
|---|------------|------------------------------|----------------|
| 0 | 2020-04-10 | 523112.121529                | 1830144.144691 |
| 1 | 2020-04-11 | 531307.744970                | 1970719.965772 |
| 2 | 2020-04-12 | 539503.368410                | 2120361.783051 |
| 3 | 2020-04-13 | 547698.991851                | 2279531.430232 |
| 4 | 2020-04-14 | 555894.615292                | 2448708.165547 |

- Tabel perbandingan Linier Regression dan SVM

### D. AR MODEL (TIME SERIES)

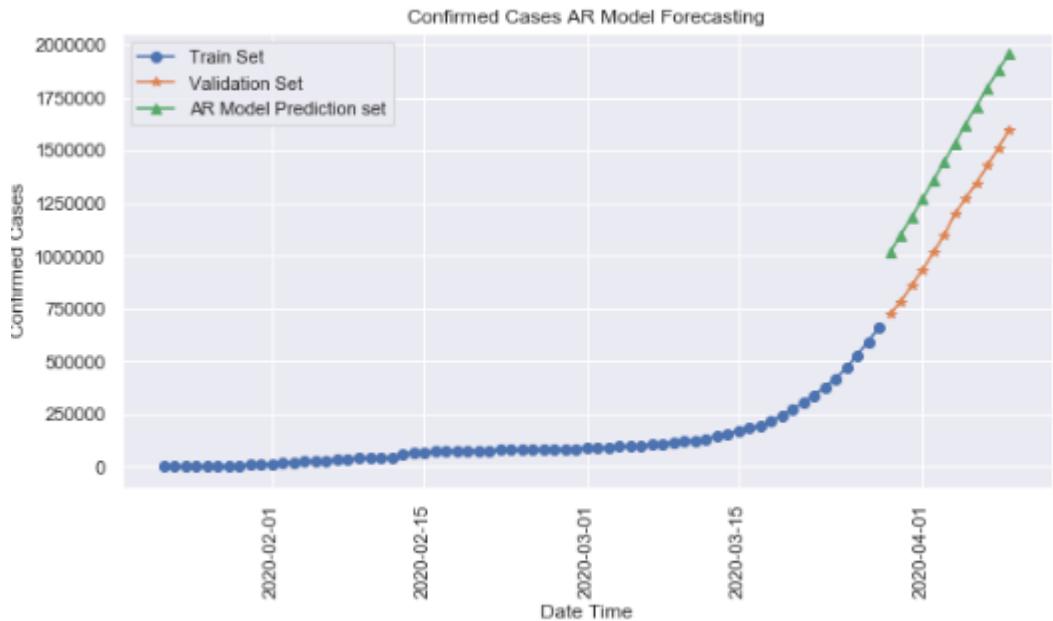
```
1 model_ar=ARIMA(log_series, (5,2,0))
2 model_ar_fit=model_ar.fit()
```

```
1 prediction_ar=model_ar_fit.forecast(len(valid))[0]
2 y_pred["AR Model Prediction"]=list(np.exp(prediction_ar))
```

- Membuat model AR (Auto Regressive) dan prediksi untuk kasus confirmed.

```
1 model_scores.append(np.sqrt(mean_squared_error(list(valid["Confirmed"]),np.exp(prediction_ar))))
2 print("Root Mean Square Error for AR Model: ",np.sqrt(mean_squared_error(list(valid["Confirmed"]),np.exp(prediction_ar))))
```

- RMSE AR Model **343519.5234643571**.



- **Grafik train kasus confirmed, validasi dan prediksi AR Model. Prediksi AR model meleset dari data actual. Bisa dilihat pada grafik diatas.**

```

1 AR_model_new_prediction=[]
2 for i in range(1,18):
3 AR_model_new_prediction.append(np.exp(model_ar_fit.forecast(len(valid)+i)[0][-1]))

```

- Prediksi Kasus kematian (Deaths) untuk AR Model.

```
1 pd.DataFrame(zip(new_date,AR_model_new_prediction),columns=["Deaths","AR Model Forecast"]).head()
```

- Membuat Tabel prediksi

|   | Deaths     | AR Model Forecast |
|---|------------|-------------------|
| 0 | 2020-04-10 | 2034664.211055    |
| 1 | 2020-04-11 | 2105945.125221    |
| 2 | 2020-04-12 | 2171511.338587    |
| 3 | 2020-04-13 | 2231625.249655    |
| 4 | 2020-04-14 | 2284718.767932    |

- Prediksi AR Model untuk kasus kematian (Deaths)

## E. ARIMA (TIME SERIES)

```
1 model_train=datewise.iloc[:int(datewise.shape[0]*0.90)]
2 valid=datewise.iloc[int(datewise.shape[0]*0.90):]
3 log_series=np.log(model_train["Confirmed"])
4 y_pred=valid.copy()
```

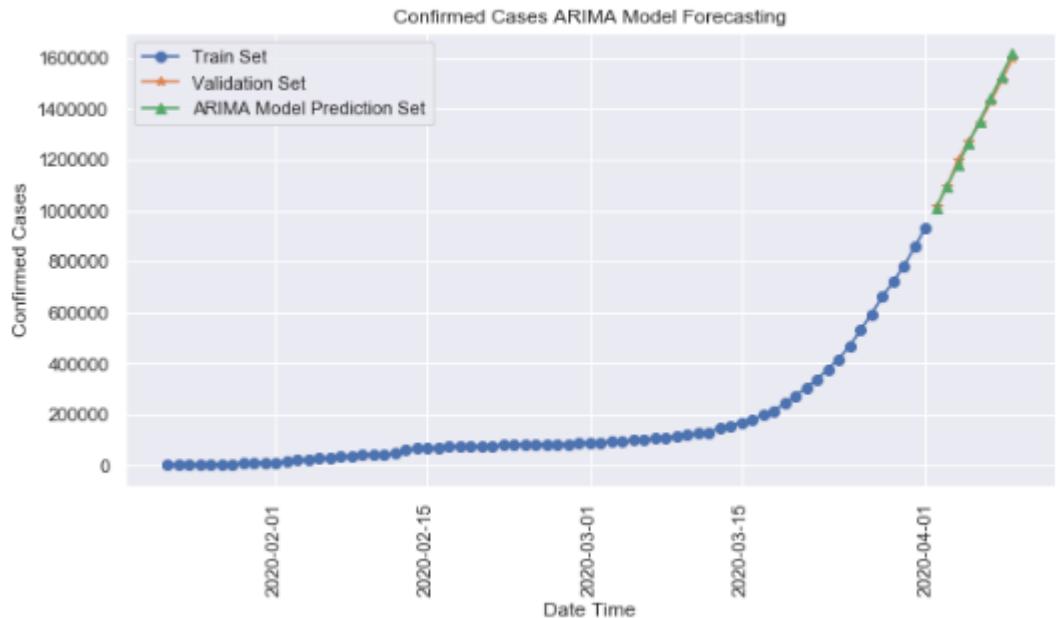
```
1 model_arima=ARIMA(log_series, (5,2,1))
2 model_arima_fit=model_arima.fit()
```

```
1 prediction_arima=model_arima_fit.forecast(len(valid))[0]
2 y_pred["ARIMA Model Prediction"]=list(np.exp(prediction_arima))
```

- Membuat model ARIMA Model untuk prediksi kasus confirmed

```
1 model_scores.append(np.sqrt(mean_squared_error(list(valid["Confirmed"]),np.exp(prediction_arima))))
2 print("Root Mean Square Error for ARIMA Model: ",np.sqrt(mean_squared_error(list(valid["Confirmed"]),np.exp(prediction_arima))))
```

- RMSE dari ARIMA Model **13097.4998404158**, Angka ini merupakan angka yang sangat baik untuk prediksi kasus confirmed.



- Grafik train kasus confirmed, validasi dan prediksi ARIMA Model. Prediksi dari ARIMA Model ini sangat akurat, bisa dilihat grafik untuk prediksi dan validasi sangat sesuai.

```

1 model_arima_deaths=ARIMA(log_deaths, (5,2,1))
2 model_arima_deaths_fit=model_arima_deaths.fit()

1 predictions_deaths=np.exp(model_arima_deaths_fit.forecast(len(valid))[0])
2 y_pred["ARIMA Death Prediction"]=predictions_deaths

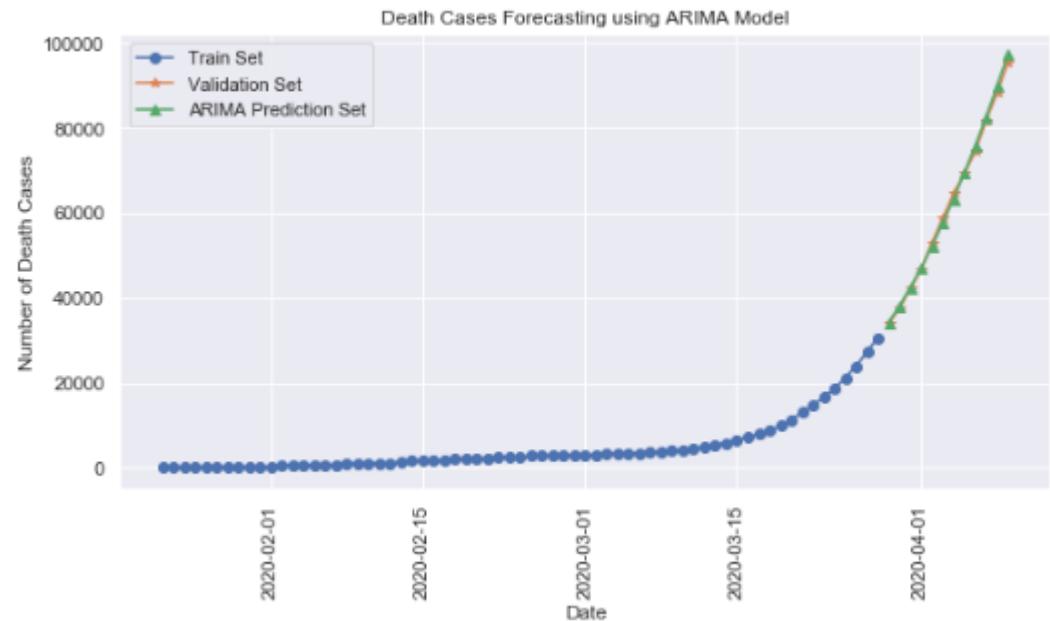
```

- Prediksi kasus kematian (Deaths) untuk model ARIMA.

```
1 print("Root Mean Square Error: ",np.sqrt(mean_squared_error(valid["Deaths"],predictions_deaths)))
```

- RMSE dari ARIMA Model untuk kasus kematian (Deaths)

**1021.8722996905492**



- Grafik prediksi, akurasi dari ARIMA Model sangat baik untuk memprediksi kasus Kematian (Deaths).

## F. PERBANDINGAN AR dan ARIMA Model kasus Kematian (Deaths)

```
1 pd.set_option('display.float_format', lambda x: '%.6f' % x)
2 model_predictions=pd.DataFrame(zip(new_date,AR_model_new_prediction, ARIMA_model_death_forecast),columns=["Dates",""
3 model_predictions.head() |
```

- Menampilkan table hasil perbandingan prediksi

|   | Dates      | AR Model       | ARIMA Model   |
|---|------------|----------------|---------------|
| 0 | 2020-04-10 | 2034664.211055 | 105060.957400 |
| 1 | 2020-04-11 | 2105945.125221 | 113289.193000 |
| 2 | 2020-04-12 | 2171511.338587 | 121401.255724 |
| 3 | 2020-04-13 | 2231625.249655 | 129805.590826 |
| 4 | 2020-04-14 | 2284718.767932 | 138607.494010 |

- Tabel Perbandingan prediksi kasus kematian untuk AR dan ARIMA Model. Pertanggal 10-14 April, sehari setelah coding dibuat (9 April proses pembuatan koding)

|    | Dates      | AR Model       | ARIMA Model   |
|----|------------|----------------|---------------|
| 12 | 2020-04-22 | 2420825.238458 | 207500.998919 |
| 13 | 2020-04-23 | 2398915.087599 | 215321.499682 |
| 14 | 2020-04-24 | 2368655.580766 | 222468.284907 |
| 15 | 2020-04-25 | 2330336.312902 | 229518.606221 |
| 16 | 2020-04-26 | 2284294.715128 | 235901.586434 |

- Tabel Perbandingan prediksi kasus kematian untuk AR dan ARIMA Model. Pertanggal 22-26 April 2020.

## G. KESIMPULAN

ARIMA Model dan SVM menunjukkan hasil paling baik untuk kasus confirmed , sedangkan untuk kasus kematian (Deaths), ARIMA menunjukkan hasil yang baik juga. Kesimpulannya, ARIMA Model adalah Machine Learning yang paling cocok untuk kasus Forecasting COVID-19 menggunakan dataset diatas. RMSE ARIMA untuk kasus confirmed mencapai **13097.4998404158**, AR Model **343519.5234643571**. sedangkan SVM mencapai angka **50167.10314784511**. Untuk kasus kematian, ARIMA berhasuk mencapai RMSE **1021.8722996905492**.