

LAPORAN
UJIAN AKHIR SEMESTER TI4141 ANALITIKA DATA



Ritzke Aisyarah

13417053

PROGRAM STUDI TEKNIK INDUSTRI
INSTITUT TEKNOLOGI BANDUNG
2020

"Saya mengerjakan UAS ini secara individu, tanpa menerima bantuan dari mahasiswa TI4141 ataupun individu lainnya. Saya memahami bahwa melakukan tindak plagiarisme dan kerja sama dalam pengerjaan UAS ini adalah terlarang, dan jika saya terbukti melakukan upaya plagiarisme dan kerja sama, saya bersedia menerima sanksi sesuai dengan kebijakan ITB."

Analysis Using CRISP-DM

1. Business Understanding

Pertumbuhan kasus baru virus COVID-19 di dunia masih terus meningkat. Hingga 30 April 2020 kasus COVID-19 di Indonesia belum memasuki masa puncak dari pandemi ini. Kebijakan pemerintah dan kedisiplinan masyarakat merupakan kombinasi yang penting dalam penanganan COVID-19. Per 30 April 2020 kebijakan yang diterapkan oleh pemerintah adalah pemberlakuan pembatasan sosial berskala besar atau yang lebih dikenal dengan PSBB. Namun tampaknya kebijakan ini belum efektif di beberapa daerah karena menurut pemantauan dari penulis yang sekarang sedang berada di daerah, masyarakat tetap melaksanakan aktivitas diluar rumah, tidak *aware* terhadap isu COVID-19, dan belum disiplin melaksanakan maklumat pemerintah.

Tujuan dari dilaksanakannya penelitian ini adalah untuk memprediksi waktu berakhir COVID-19 di Indonesia berdasarkan dataset kasus COVID-19 yang ada di Italia.

Project dapat dikatakan berhasil jika :

- Dapat memprediksi puncak COVID-19 di Indonesia dari data Italia
- Dapat memberikan rekomendasi bagi *policy maker* dalam penanganan COVID-19 di Indonesia

2. Data Understanding

Data yang diberikan terdiri dari dua dataset yaitu *data province* dan *data regional*. Data tersebut termasuk kedalamnya variabel kasus penambahan per hari, kasus kesembuhan per hari, kasus kematian per hari, dan total kasus dari setiap *region* di Italia. Penulis juga mengumpulkan data kasus di Indonesia.

3. Data Preparation

Dari dua dataset yang diberikan, hanya terdapat *missing data* pada *data province* yaitu pada variabel *province geocode* terdapat 60 data yang bersifat nul. Untuk *data region* tidak terdapat *missing data*.

Data preparation yang dilakukan :

- Memisahkan data per masing-masing *region code* 1 – 20 untuk dicari tau *region* mana yang memiliki trend yang mirip dengan kasus di Indonesia
- Mengelompokkan data kejadian per tanggal sehingga diperoleh rekapitulasi total kasus, kesembuhan, dan kematian per tanggal.

Dalam mencari *region* yang memiliki trend yang mirip dengan Indonesia, pertama dicari rata-rata penambahan kasus perhari selama 60 hari yang mirip dengan Indonesia, dari data Indonesia didapat rata-rata penambahan kasus Indonesia sebagai berikut :

```
data_indonesia["Kasus Baru"].mean()
```

```
Out[29]: 183.96363636363637
```

Kemudian dicari rata-rata penambahan kasus di Itali per masing-masing region sebagai berikut :

```
total1_total_cases1["New_Actually_Positive"].mean()
Out[7]: 385.6666666666667
total2_total_cases2["New_Actually_Positive"].mean()
Out[8]: 18.266666666666666
total3_total_cases3["New_Actually_Positive"].mean()
Out[9]: 1169.3166666666666
total4_total_cases4["New_Actually_Positive"].mean()
Out[10]: 102.7
total5_total_cases5["New_Actually_Positive"].mean()
Out[11]: 281.3333333333333
total6_total_cases6["New_Actually_Positive"].mean()
Out[12]: 47.63333333333333
total7_total_cases7["New_Actually_Positive"].mean()
Out[13]: 117.48333333333333
total8_total_cases8["New_Actually_Positive"].mean()
Out[14]: 395.3833333333333
total9_total_cases9["New_Actually_Positive"].mean()
Out[15]: 146.33333333333334
total10_total_cases10["New_Actually_Positive"].mean()
Out[16]: 22.7
total11_total_cases11["New_Actually_Positive"].mean()
Out[17]: 99.2
total12_total_cases12["New_Actually_Positive"].mean()
Out[18]: 100.88333333333334
total13_total_cases13["New_Actually_Positive"].mean()
Out[19]: 46.416666666666664
total14_total_cases14["New_Actually_Positive"].mean()
Out[20]: 4.733333333333333
total15_total_cases15["New_Actually_Positive"].mean()
Out[21]: 70.63333333333334
total16_total_cases16["New_Actually_Positive"].mean()
Out[22]: 63.98333333333334
total17_total_cases17["New_Actually_Positive"].mean()
Out[23]: 5.933333333333334
total18_total_cases18["New_Actually_Positive"].mean()
Out[24]: 17.816666666666666
```

```
total19_total_cases19["New_Actually_Positive"].mean()
```

```
Out[25]: 48.766666666666666
```

```
total20_total_cases20["New_Actually_Positive"].mean()
```

```
Out[26]: 20.9
```

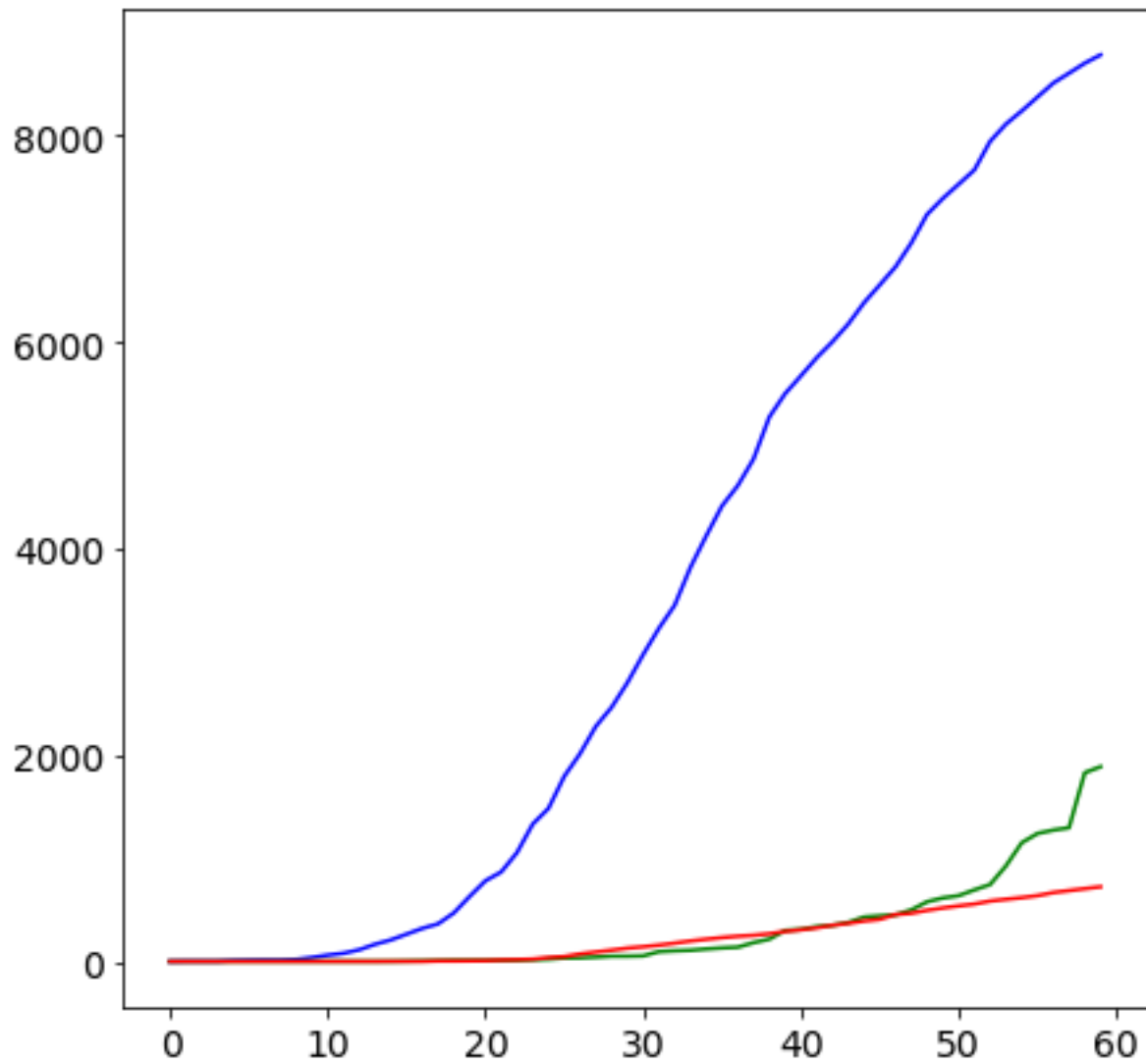
Sedangkan rata-rata penambahan kasus untuk Itali sebagai berikut :

```
tc1['New_Actually_Positive'].mean()
```

```
Out[35]: 3166.0833333333335
```

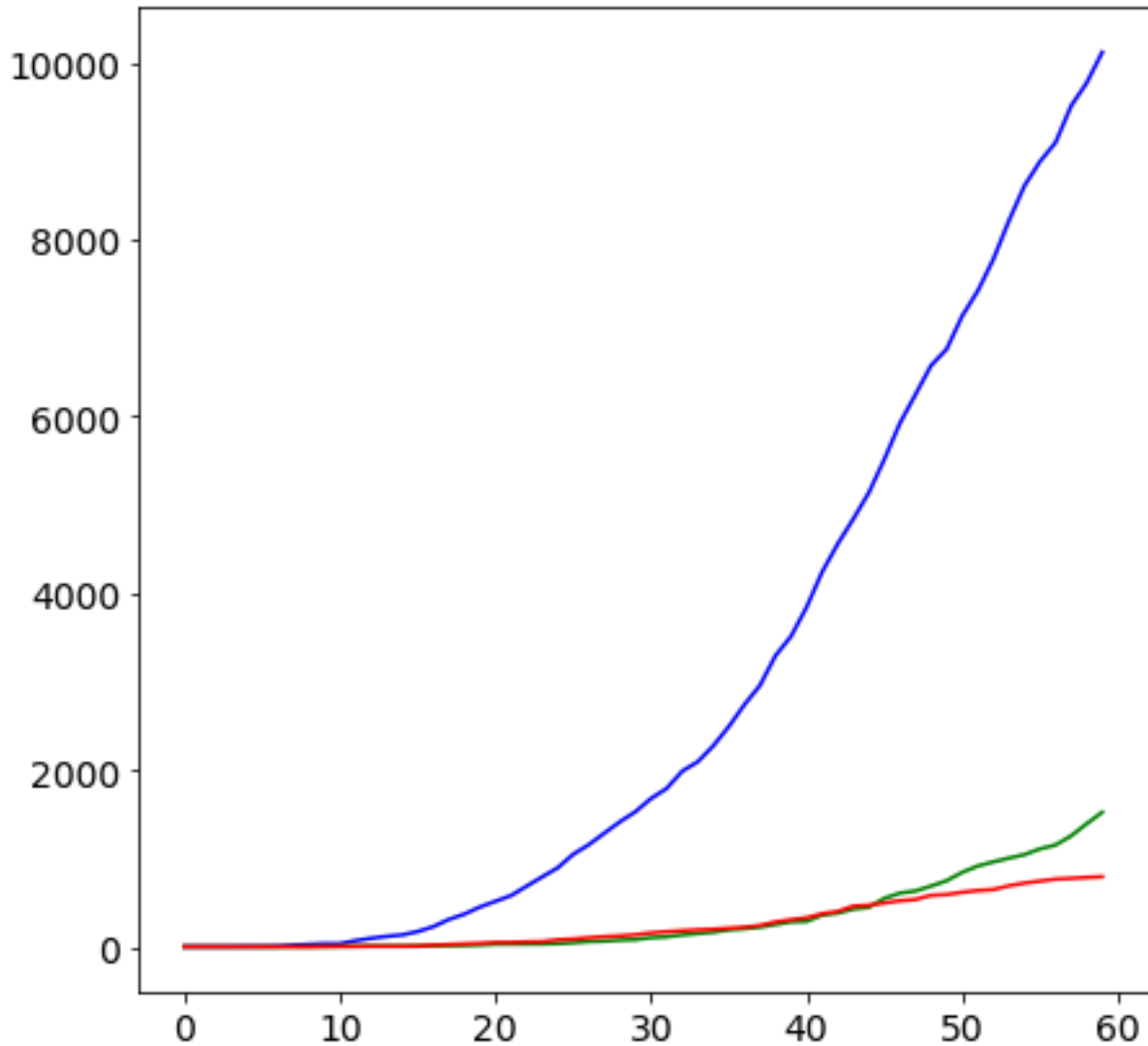
Karena penambahan kasus untuk Italia memiliki gaalt yang jauh dengan Indonesia, maka dicari region yang memiliki trend kasus yang sama dengan Indonesia yaitu kasus untuk region no **9**.

Berikut grafik total kasus, kesembuhan, dan kematian untuk Italia region 9



Gambar 1 Trend Itali Region 9

Berikut grafik total kasus, kesembuhan, dan kematian untuk Indonesia

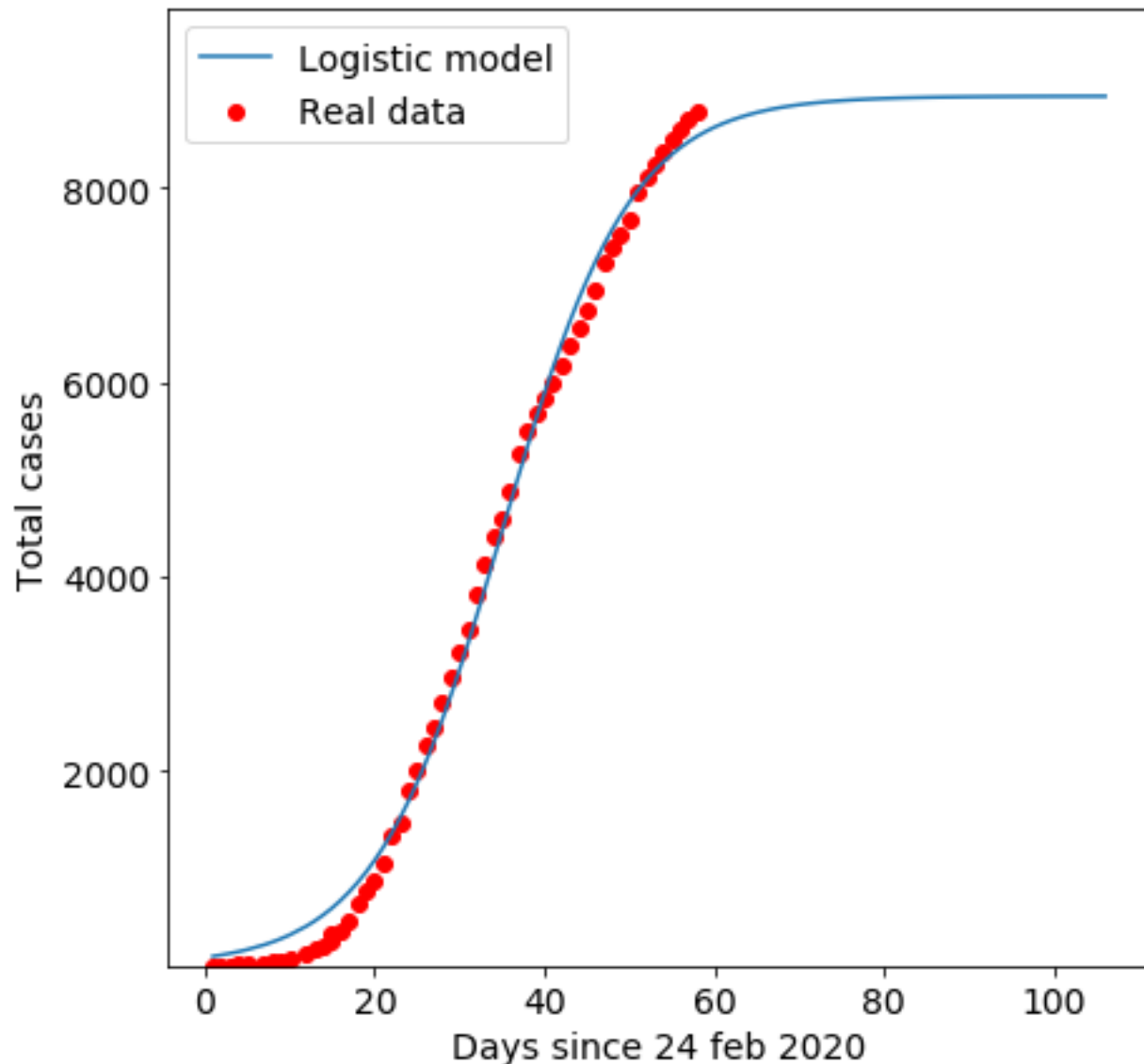


Gambar 2 Trend Indonesia

4. Modelling

Dalam tahap pemodelan data, digunakan **logistic regression** untuk memprediksi puncak pandemi COVID-19 di Indonesia dari data Italia region 9. Alasan digunakannya model *logistic regression* adalah untuk mengetahui puncak dari masa pandemi karena akhir masa COVID-19 di Indonesia berdasarkan prediksi dari *Singapore University Technology and Design* (SUTD) mundur menjadi 7 Oktober 2020 dari prediksi awal September 2020 (Franedya, 2020). Sehingga, diperlukan informasi terkait puncak COVID-19 di Indonesia agar masyarakat Indonesia tetap *aware* dan tetap melakukan upaya agar penularan COVID-19 dapat diminimalisir karena kasus penambahan masih akan terus meningkat hingga masa puncak.

Berikut grafik yang menampilkan hasil dari model *logistic regression* di Italia region 9.



Gambar 3 Logistic Model Italia Region 9

Dari model tersebut didapat bahwa puncak COVID-19 terjadi setelah 107 hari sejak kasus pertama. Sehingga jika dicocokkan dengan kejadian di Indonesia yang mana kasus pertama terhitung pada 2 Maret 2020, 107 hari setelah itu jatuh pada tanggal 17 Juni 2020. Sehingga berdasarkan model *logistic regression* diprediksi puncak COVID-19 jatuh pada tanggal **17 Juni 2020** sehingga laju penambahan kasus akan terus positif hingga masa puncak pandemi.

5. Evaluation

Berdasarkan hasil prediksi dari SUTD, akhir pandemi COVID-19 di Indoneis jatuh pada 7 Oktober 2020, sehingga hasil prediksi masa puncak yang diperoleh, 17 Juni 2020 dapat diterima dan sesuai dengan hasil prediksi oleh SUTD

Model yang digunakan sederhana dan hanya dapat memprediksi waktu puncak pandemi COVID-

19 di Indonesia belum dapat memprediksi waktu berakhir COVID-19. Untuk penelitian berikutnya, data juga dapat digunakan untuk memprediksi akhir COVID-19 bisa dengan data Indonesia maupun penyesuaian dari data negara lain dengan model-model lainnya seperti SVM, ARIMA, dan lain sebagainya.

6. Deployment

Pada tahap *deployment*, dapat dilakukan dengan mengintegrasikan model hasil prediksi pada kanal resmi Kawal COVID-19 di Indonesia. Hal ini bertujuan agar masyarakat *aware* terhadap pandemi ini dan taat melaksanakan maklumat pemerintah untuk tetap melakukan gerakan dirumahaja. Selain itu, dengan mengintegrasikan hasil model pada kanal resmi pemerintah, dapat memberi semangat positif kepada masyarakat bahwa bangsa Indonesia akan sampai pada akhir COVID-19 ini dengan terlebih dahulu bersiap-siap untuk menghadapi masa puncak.

Kesimpulan dan saran untuk *stakeholder* Indonesia terkait penyebaran COVID-19 :

1. Berdasarkan model yang diprediksi, Indonesia belum memasuki masa puncak COVID-19. Sehingga diharapkan pemerintah dapat lebih tegas terkait protokol untuk melaksanakan *physical distancing*.
2. Kepada Pemerintah Indonesia, diharapkan mempunyai sanksi tegas bagi siapa saja, baik untuk TIM SATGAS, TIM Medis, dan masyarakat yang tidak menjalankan protokol masing-masing dengan benar.
3. Kepada Pemerintah Indonesia, diharapkan terus dapat meningkatkan fasilitas pelayanan kesehatan di Indonesia, karena tidak dapat dibandingkan terkait fasilitas kesehatan Indonesia dengan negara maju seperti Italia. Untuk Italia pun mereka masih kewalahan dalam penanganan COVID-19. Diharapkan Indonesia terus meningkatkan fasilitas pelayanan kesehatan dengan memberi alokasi yang lebih kepada fasilitas kesehatan. Karena jika tidak, maka dapat lebih buruk dari Italia. Selain itu, peningkatan fasilitas kesehatan dapat membuat jumlah tenaga medis yang bekerja lebih banyak, karena pada saat sekarang ini banyak tenaga medis yang tidak turun lapangan karena fasilitas yang tidak memadai.
4. Kepada masyarakat Indonesia, diharapkan patuh dan taat kepada maklumat pemerintah untuk melaksanakan *physical distancing*, menghindari kerumunan, dan menjaga kebersihan dan kesehatan diri pribadi.