**Prediksi Kasus Kasus Positif dan Kematian di Indonesia dengan Model Regressi**

Laporan ini disusun untuk memenuhi tugas mata kuliah

Analisa Data dan Bisnis

Semester II Tahun Akademik 2019/2020



**Oleh:**

23519027 Suwardiman

**PROGRAM STUDI TEKNIK INFORMATIKA**

**INSTITUT TEKNOLOGI BANDUNG**

**BANDUNG**

**2020**

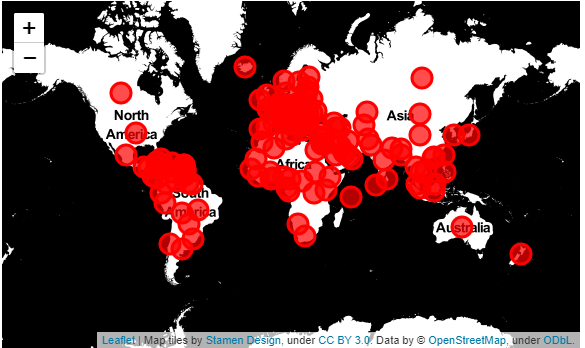
1. **Latar Belakang**

Dalam istilah sederhana, dilansir dari The Sun, Covid-19 adalah singkatan dari Corona (CO), Virus (VI) Disease (D) dan tahun 2019 (19), yang mana virus corona Covid-19 ini pertama kali muncul di tahun 2019. Organisasi Kesehatan Dunia (WHO) pun akhirnya menetapkan Covid-19 untuk menyebut virus corona yang sedang mewabah di seluruh dunia ini.

Virus corona jenis baru yang disebut SARS-CoV-2, yang sebelumnya disebut 2019-nCoV adalah virus jenis baru yang belum diidentifikasi pada manusia sebelumnya.

Penyakit yang disebabkan oleh SARS-CoV-2 inilah pada akhirnya disebut Covid-19. Jadi, virus jenis baru ini akhirnya lebih dikenal sebagai corona Covid-19.

Coronavirus adalah keluarga besar virus yang bisa menyebabkan penyakit, mulai dari flu biasa hingga penyakit pernapasan paling parah, seperti Sindrom Pernapasan Timur Tengah (MERS) dan Sindrom Pernapasan Akut Parah (SARS).



Gambar 1. Peta Persebaran Covid-19

Sejak pertama kali virus ini terdeteksi di Wuhan, China, pada Desember 2019, wabah ini telah berkembang sangat cepat. WHO lalu melabeli wabah virus corona Covid-19 ini sebagai pandemi global.

Gejala khas corona Covid-19 sendiri termasuk demam, batuk, kesulitan bernapas, nyeri otot hingga kelelahan. Pada kasus yang lebih parah, virus ini bisa menyebabkan pneumonia berat, sindrom gangguan pernapasan akut, sepsis dan syok septik.

1. **Problem.**

Berkaca pada negara Italia yang menurut berita dan wawancara yang dilakukan kepada warga negara Indonesia yang terjebak di Italia, ketidakmampuan Italia memperkirakan pasien Covid-19 menyebabkan ketidaksiapan fasilitas kesehatan serta tenaga medis yang ada sehingga banyak sekali rumah sakit yang membludak oleh pasien, bahkan pasien-pasien yang sudah dinyatakan positif-pun harus *self-isolation* di rumah masing-masing karena tidak ada space lagi di rumah sakit.

Jumlah kasus kematian yang tinggi pun menjadi masalah di negara tersebut, dikatakan bahwa sekarang sangat kesulitan mendapatkan peti jenazah dan lahan untuk pasien yang dinyatakan meninggal dunia.

Berdasarkan hal yang sudah dijabarkan sebelumnya akan dibuat sebuah model yang mampu melakukan prediksi terhadap berapa jumlah kasus positif dan kasus kematian di Indonesia untuk hari yang akan datang?

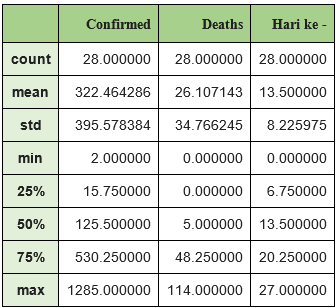
Hal ini penting dilakukan untuk dapat mempersiapkan diri terhadap kemungkinan-kemungkinan yang terjadi sehingga masalah yang sudah diprediksi dapat tertangani dengan baik. Dalam hal ini prediksi kasus positif dan kasus kematian oleh Covid-19.

Dengan mengetahui prediksi jumlah ini pemerintah dapat mempersiapkan diri, baik dari tenaga medis serta fasilitas kesehatan yang memenuhi terhadap prediksi yang diberikan.

1. **Dataset**

Dataset didapatkan dari kaggle : <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>. Kemudian diambil data untuk Indonesia, serta fitur-fitur yang digunakan yaitu confirmed, deaths dan hari sejak kasus pertama ditemukan di Indonesia

Deskripsi data:



1. **Metode yang digunakan**

* Linear Regression

Digunakan untuk melakukan prediksi jumlah kasus positif. Prediksi yang dilakukan dengan menjadikan hari ke-i sejak kasus pertama ditemukan hingga data terbaru sebagai independen variabel (X) dan jumlah kasus positif sebagai dependen variabel (Y).

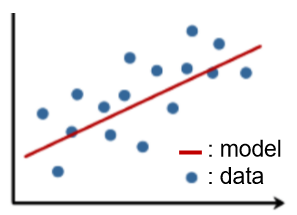
Regresi linear dapat dirumuskan sebagai berikut :

𝑦(𝒙) = 𝑎0 + 𝑎1𝑥1 + 𝑎2𝑥2 + ⋯+ 𝑎𝑛𝑥𝑛

di mana: 𝑎𝑖 = koefisien regresi

𝑥𝑖 = atribut / fitur / var. independen

𝑦 = output prediksi / var. dependen



Gambar 1. Gambar Linear Regression

* Polynomial Regression.

Digunakan untuk melakukan prediksi jumlah kasus kematian. Prediksi dilakukan dengan menjadikan kasus positif sebagai independen variabel (X) dan kasus kematian sebagai dependen variabel (Y). Model dibuat dengan Polynomial derajat 5.

Polynomial dapat dirumuskan :

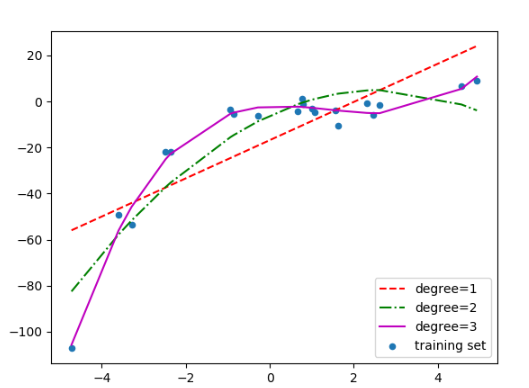
𝑦 (𝒙)= a0+a1x1+a2x12+ ….. + anx1n

di mana:

𝑎𝑖 = koefisien regresi

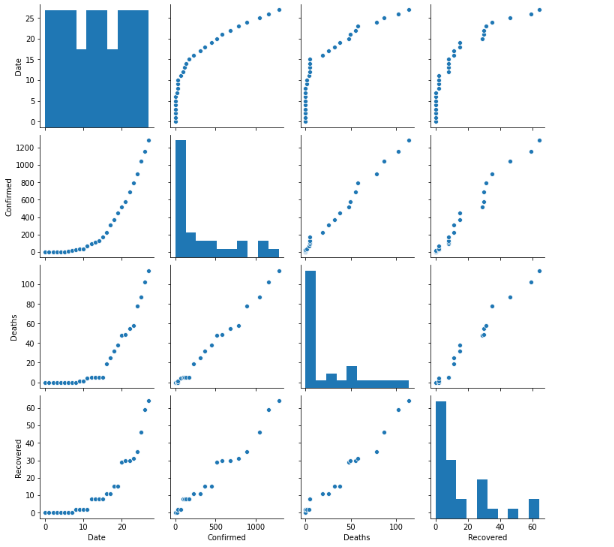
𝑥𝑖 = atribut / fitur / var. independen

𝑦 = output prediksi / var. dependen

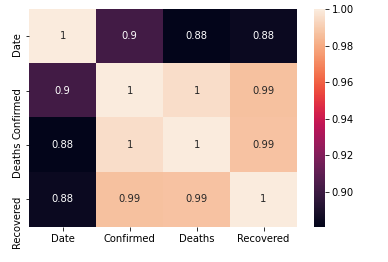


Gambar 2. Contoh Grafik Regresi polynomial

Kedua metode dipilih dengan melihat hasil eksplorasi berdasarkan nilai korelasi dan scatter-plot yang ada pada Gambar 3 dan Gambar 4.



Gambar 3. Gambar Scatter Plot



Gambar 4. Gambar Heatmap Pearson Correlation

Terlihat hubungan antara Confirmed & Deaths cenderung linier dan hubungan antara date(index hari) & Confirmed cenderung polynomial. Model akan dibangun dengan 34 data dengan pembagian 90% training dan 10% testing.

1. **Insight dan Hasil**

Setelah melakukan training dan testing pada 34 data dengan perbandingan 90% training dan 10% testing didapatkan hasil evaluasi model sebagai berikut :

Tabel 1. Tabel Evaluasi Model

|  |  |  |
| --- | --- | --- |
| Metode | RMSE Training | RMSE Testing |
| Linear Regression(death) | 4.155 | 2.165 |
| Polynomial Regression (Confirmed) | 12.88 | 15.58 |

Tabel 2. Tabel prediksi model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Confirmed Prediction | Deaths Prediction | Actual Confirmed | Actual Deaths |
| 3/28/2020 | 1144.94258 | 101.281 | 1155 | 102 |
| 3/29/2020 | 1271.13888 | 112.778 | 1285 | 114 |
| 3/30/2020 | 1402.32778 | 124.729 | 1414 | 122 |
| 3/31/2020 | 1537.9429 | 137.084 | 1528 | 136 |
| 4/1/2020 | 1677.36482 | 149.785 | 1677 | 157 |
| 4/2/2020 | 1819.92258 | 162.772 | 1790 | 170 |
| 4/3/2020 | 1964.89526 | 175.979 | 1986 | 181 |
| 4/4/2020 | 2111.51345 | 189.337 | 2092 | 191 |
| 4/5/2020 | 2258.96076 | 202.769 | 2273 | 198 |
| 4/6/2020 | 2406.3754 | 216.199 | 2491 | 209 |
| 4/7/2020 | 2552.851629 | 229.543 | 0 | 0 |

Tabel 3. Tabel evaluasi hasil prediksi tanggal 28 Maret-6 April

|  |  |
| --- | --- |
| Metode | RMSE |
| Linear Regression(death) | 4.29 |
| Polynomial Regression (Confirmed) | 16.49 |

Berdasarkan hasil prediksi yang dilakukan oleh model dari tanggal 28 maret – 6 april 2020 hasil RMSE terlihat tidak jauh berbeda dengan hasil RMSE pada training dan testing yaitu sebesar 4,29 untuk prediksi kasus meninggal dan 16,49 untuk prediksi kasus positif.

1. **Diskusi**

Rencana sebelumnya mau menggunakan data RS rujukan untuk melihat bagaimana korelasi antara jumlah RS rujukan yang disediakan pemerintah dengan jumlah kematian yang ada.

Terdapat kesulitan mendapatkan data, data yang didapatkan terakhir tidak berupa time series, hanya berupa data yang terus terupdate setiap harinya seperti pada situs-situs berita yang ada, sehingga tidak ada sisi historis/perkembangan untuk dapat melihat perubahan setiap harinya yang dapat dibandingkan dengan jumlah kasus kematian setiap harinya.