

Menelusuri COVID-19 di Dunia dan di Indonesia dengan Model Regresi SVM, Bayesian dan Gaussian

Hanna Arini Parhusip¹⁾

¹⁾Program Studi Matematika Fakultas Sains dan Matematika
Universitas Kristen Satya Wacana
hanna.parhusip@uksw.edu

(Article History: Received 03-04-2020; Accepted 14-05-2020; Published 03-06-2020)

ABSTRAK

Pada artikel ini ditunjukkan penelusuran data total kasus COVID-19 di dunia dan di Indonesia. Penelitian ini bertujuan memprediksi banyaknya kasus COVID-19 terutama di Indonesia pada waktu tertentu mendatang. Metode yang digunakan adalah metode SVM (*Support Vector Machine*) dan regresi Bayesian Ridge dimana kedua metode berhasil memprediksi untuk total kasus COVID-19 di beberapa negara di dunia dan hasil Bayesian Ridge lebih baik. Untuk itulah kedua metode ini digunakan untuk menelusuri data total kasus COVID-19 di Indonesia. Akan tetapi pengaruh *social distancing* belum ditunjukkan sehingga digunakan model eksponensial dan Gaussian. Model eksponensial digunakan sebagai nilai awal untuk prediksi pada model Gaussian. Pada model Gaussian diperoleh prediksi dalam bentuk 2 skenario yaitu skenario terburuk dan skenario realistis. Skenario terburuk memberikan total kasus pada 24 April 2020 sebanyak 14675. Hal ini tidak realistis karena data terakhir pada 31 Maret 2020 sebesar 1677. Oleh karena itu diusulkan skenario realistis dimana total kasus puncak berlaku pada 24 April 2020 dengan total kasus 4242. Akan tetapi ternyata data total kasus positif COVID-19 pada 24 April 2020 sebanyak 8211. Dengan memperhatikan hasil prediksi dan yang diperoleh dari data, pendekatan yang sudah dilakukan masih memerlukan perbaikan.

Kata Kunci: *Bayesian Ridge*; COVID-19; *Gaussian*; *Social distancing*; SVM (*Support Vector Machine*)

Study on COVID-19 in the World and Indonesia Using Regression Model of SVM, Bayesian Ridge and Gaussian

ABSTRACT

In this article, research on data of total cases COVID-19 in the world and Indonesia is presented. The purposes of this research is to predict the number of cases COVID-19 mainly in Indonesia in the near future. The used methods are regression of Support Vector Machine (SVM) and Bayesian Ridge predicting 10 days forward in several countries where Bayesian Ridge gives better result. Similarly, both methods are employed to study the total cases COVID-19 in Indonesia. To include social distancing in the study, an exponential model is applied for initialization and then Gaussian is used to predict future total cases. Using Gaussian model, one has 2 scenarios of future predictions. The number of total cases on 24th April 2020 is 14675 and considerable not realistic. This is updated i.e. the total number of cases in 24 April 2020 is 4242. Actually, the number of total positive cases COVID-19 on 24th April 2020 was 8211. Compared to the prediction with the real data, the approach of this research should be improved to give better prediction.

Keywords: *Bayesian Ridge*; COVID-19; *Gaussian*; *Social distancing*; SVM (*Support Vector Machine*)

PENDAHULUAN

Telah diketahui bahwa munculnya virus COVID-19 diumumkan pertama kali di

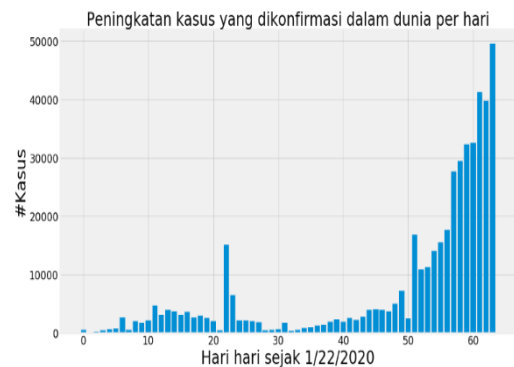
Wuhan pada Desember 2019. Laju penularan yang sangat cepat oleh karena mobilitas masyarakat dunia yang tinggi menyebabkan kesiapan negara-negara di dunia dipandang

tidak cukup dalam mengantisipasi penularan yang terjadi. Adanya pendataan yang cepat dari John Hopkins University membantu untuk mencermati penyebaran virus ini pada berbagai negara di dunia. Beberapa lembaga statistik di dunia juga melakukan pendataan yang selalu diperbaharui, misalkan (Ortiz-Ospina, 2020). Demikian pula kepedulian sosial muncul atas pengaruh yang ditimbulkan akibat perkembangan laju COVID-19 yang meningkat sangat cepat (Hamid, 2020). Beberapa pendataan kemudian diilustrasikan pada grafik sehingga dapat dicermati sebagaimana ditunjukkan pada Gambar 1 dimana *World Health Organization* (WHO) mencatat kejadian COVID-19 sejak 22 Januari 2020 hingga 16 Maret 2020 di berbagai belahan dunia. Salah satu yang terlihat peningkatan kasus yang sangat tinggi selain Wuhan di China juga terjadi di Italia sebagaimana ditunjukkan pada Gambar 2. Selanjutnya diilustrasikan perbandingan total konfirmasi kasus COVID-19 Cina, Italia dan US tersebut pada Gambar 3.

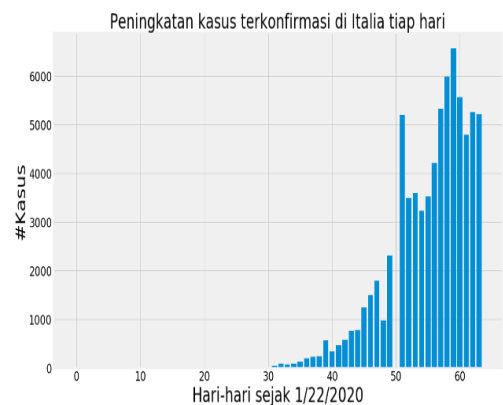
Pendataan secara cepat dan *uptodate* ini telah dilakukan dan diharapkan dapat memberikan penjelasan kepada semua pihak untuk dapat segera melakukan tindakan pencegahan dimana sekarang diberlakukan *social distancing* (SD) ataupun *physical distancing* (PD) yaitu pembatasan interaksi sosial antar manusia maupun kegiatan yang melibatkan kelompok.

Sejak 16 Maret 2020 di Indonesia SD atau PD diberlakukan. Telah direkomendasikan oleh penelitian bagaimana SD ataupun PD berpengaruh terhadap memperlambat laju kasus COVID-19 di Wuhan (Prem *et al.*, 2020) serta di seluruh dunia melakukan SD (Pambuccian, 2020) hingga 2 April 2020 (hingga artikel ini ditulis masih dalam situasi SD). Aksi SD ini telah dilakukan dan laju pertumbuhan pasien yang terkonfirmasi COVID-19 perlu dipelajari untuk melihat ada tidaknya pengaruh yang signifikan baik di Indonesia. Pada mulanya, tulisan ini menunjukkan penelusuran pendataan kasus yang dikonfirmasi pada tingkat dunia dimana data di WHO di mulai pada 22 Januari 2020 sedangkan data di Indonesia dimulai dari 6 Maret 2020. Dengan mempelajari COVID-19 di dunia, maka laju total kasus di Indonesia dapat diprediksi. Oleh karena itulah penelitian ini dilakukan, yaitu memberikan prediksi pada banyaknya kasus

COVID-19 pada waktu selanjutnya berdasarkan data yang ada.



Gambar 1. Laju peningkatan kasus per hari yang dikonfirmasi di dunia yang dicatat WHO sejak 22 Januari 2020 hingga 16 Maret 2020; banyaknya kasus (sumbu vertikal), banyaknya hari (sumbu horizontal).

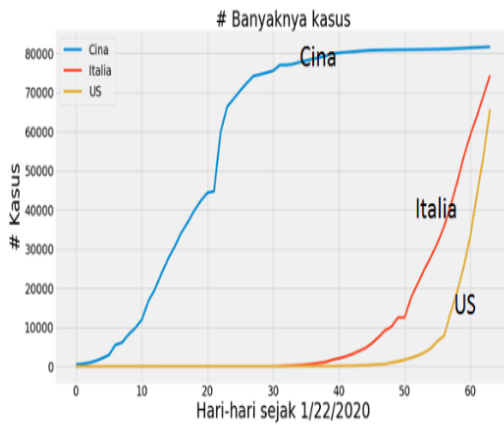


Gambar 2. Laju peningkatan kasus yang dikonfirmasi di Italia yang dicatat WHO sejak 22 Januari 2020 hingga 16 Maret 2020; banyaknya kasus (sumbu vertikal), banyaknya hari (sumbu horizontal) di dunia (atas) dan di Italia (bawah)

Metode yang digunakan adalah SVM (*Support Vector Machine*) dan regresi Bayesian Ridge. Metode tersebut biasa digunakan dalam melakukan klasifikasi yang biasa dilakukan dengan Python (Min-Wei *et al.*, 2018; Lamidi, 2018; Mckinney, 2018) sebagai metode yang banyak digunakan pada *Machine Learning* pada pengolahan big data (Demidova *et al.*, 2016).

Akan tetapi masyarakat menghendaki bahwa model yang digunakan dapat menunjukkan waktu-waktu dimana terjadi kasus COVID-19 mencapai puncak dan kapan terjadi kasus COVID-19 menurun karena adanya pengaruh SD. Untuk itulah model eksponensial juga diperkenalkan yang kemudian digunakan sebagai bagian dari

model Gaussian. Hal inilah yang menjadi novelti atas penelitian ini dimana model dapat memberikan prediksi banyaknya total kasus yang muncul pada waktu tertentu.

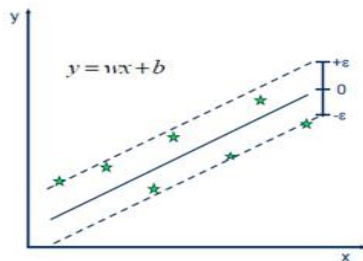


Gambar 3. Banyaknya kasus terkonfirmasi (sumbu vertikal) , banyaknya hari (sumbu horizontal) di dunia yang dicatat WHO dari 3 negara yaitu Cina, Italia, US (kanan) sejak 22 Januari hingga 16 Maret 2020.

METODE PENELITIAN

Model Regresi SVM

SVM merupakan salah satu metode klasifikasi dalam *machine learning*. Akan tetapi dapat pula untuk regresi yang disebut *Support Vector Regression* (SVR). Metode ini mencocokkan sebanyak mungkin data dengan garis serta memperhatikan margin sebesar epsilon sebagaimana ditunjukkan pada Gambar 4 untuk metode SVR linear.



Gambar 4. Ilustrasi SVM dengan margin ϵ

SVR menggunakan prinsip yang sama dengan SVM untuk klasifikasi dengan hanya sedikit perbedaan. Ide dasar SVR sama dengan SVM yaitu : meminimalkan error, pembuatan *hyperplane* yang memaksimalkan margin. Masalah optimasi pada SVM dengan kernel ϕ yang memetakan x_i ke $\phi(x_i)$ adalah:

$$\min_{\mathbf{w}, b, \zeta} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \zeta_i \tag{1}$$

dengan kendala

$$y_i(\mathbf{w} \cdot \phi(x_i) + b) \geq 1 - \zeta_i$$

untuk semua $1 \leq i \leq n, \zeta_i \geq 0$. (2)

Kernel ini merupakan fungsi yang dapat dipilih seperti polinomial ,sigmoid yang disediakan dalam paket python untuk mendefinisikan kurva pemisah data. Demikian pula parameter C pada fungsi tujuan persamaan (1) merupakan input pada fungsi yang sudah ada yang merupakan parameter regulasi pada metode SVM dimana data dapat memuat *outlier*. Pada komputasi, C dapat diatur diantara $0 < C < 1$ (Kowalczyk, 2017).

Model Regresi Bayesian

Model linear regresi Bayesian Ridge merupakan perpaduan antara regresi Ridge dan linear regresi Bayesian (Permai & Tanty, 2018; Qian, 2018). Regresi Ridge menggunakan regularisasi dengan norm L_2 sedangkan regresi Bayesian model didefinisikan secara probabilistik dengan prior eksplisit untuk parameter-parameter. Pemilihan prior berpengaruh dalam regularisasi. Jadi keduanya tidak sama karena ridge regresi adalah suatu model regresi, dan pendekatan Bayesian adalah langkah yang lebih umum untuk mendefinisikan dan mengestimasi model statistik yang dapat diaplikasikan pada model-model yang berbeda. Diketahui bahwa regresi standar dalam 1 variabel dapat ditulis sebagai

$$y = \beta_0 + \beta_1 x + \epsilon$$

Untuk mengubah regresi standar ini dalam regresi Bayesian maka regresi standar tersebut ditulis dalam model probabilistik dalam bentuk

$$\mu_i = \alpha + \beta x_i$$

$$y_i \sim \mathcal{N}(\mu_i, \sigma)$$

yang berarti variabel dependen Y mengikuti distribusi normal dengan parameter μ_i yaitu fungsi linear X yang diparameterisasi oleh α, β dan deviasi standard σ . Jika model linear dengan OLS (*Ordinary Least Square*) digunakan, probabilitas tidak perlu dipermasalahkan karena nilai optimal α, β yang dicari dengan meminimalkan kuadrat error dalam pencocokan data dengan pendekatan. Sebaliknya, model tersebut memaksimalkan fungsi *likelihood*

$$\arg \max_{\alpha, \beta, \sigma} \prod_{i=1}^n \mathcal{N}(y_i; \alpha + \beta x_i, \sigma) \tag{3}$$

dimana \mathcal{N} adalah fungsi densitas dari distribusi normal yang dihitung pada titik-titik y_i dengan parameter $\alpha + \beta x_i$ dan σ .

Pada pendekatan Bayesian, sebagai pengganti memaksimumkan fungsi likelihood, diasumsikan distribusi prior untuk parameter dan menggunakan teorema Bayesian

posterior \propto likelihood \times prior.

Fungsi likelihood sama dengan persamaan (3), tetapi yang berubah adalah bahwa distribusi prior diasumsikan untuk parameter-parameter α, β, σ yang diduga dan melibatkannya dalam persamaan

$$f(\alpha, \beta, \sigma | Y, X) \propto$$

$$\prod_{i=1}^n \mathcal{N}(y_i; \alpha\beta x_i, \sigma) f_\alpha(\alpha) f_\beta(\beta) f_\sigma(\sigma).$$

Penyusunan MLE dan MAP

Terdapat 2 error yang digunakan yaitu MLE dan MAP. MLE (*Maximum Likelihood Estimation*) dan MAP (*Maximum a posterior probability*) dimana MLE memaksimumkan keserupaan data yang ditentukan parameter dan MAP memilih parameter yang memaksimumkan probabilitas posterior berdasarkan data. Sebutlah himpunan pasangan data $D = \{x_i, y_i\}$ dimana $x_i \in \mathcal{R}^n$ dan $y_i \in \mathcal{R}$. Model linear regresi menduga nilai y_i sebagai kombinasi linear dari variabel x_i dimana

$$\begin{aligned} y_i &= w_0 + \sum_j w_j x_{ij} \\ &= w_0 + \mathbf{w}^T \mathbf{x}_i \end{aligned} \quad (4)$$

Selanjutnya, tambahkan 1 sebagai vektor pada x_i dan menggabungkan w_0 dan w_i pada satu vektor, persamaan (4) dapat ditulis dalam bentuk

$$\mathbf{y}_i = \mathbf{w}^T \mathbf{x}_i.$$

Nilai observasi \mathbf{y}_i diasumsikan mempunyai derau Gaussian, sehingga ditulis

$$\mathbf{y}_i = \mathbf{w}^T \mathbf{x}_i + \epsilon \text{ dimana } \epsilon \sim \mathcal{N}(0, \sigma^2).$$

Fungsi *likelihood* untuk kasus ini adalah

$$\begin{aligned} \mathcal{L}(D | \mathbf{w}, \sigma) \\ = (2\pi\sigma^2)^{-n/2} \prod_i \exp\left[-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right]. \end{aligned}$$

Sehingga logaritma \mathcal{L} adalah

$$\begin{aligned} \ln \mathcal{L} &= -\frac{n}{2} \ln(2\pi\sigma^2) \\ &\quad - \frac{1}{2\sigma^2} \sum_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2. \end{aligned}$$

Berdasarkan MLE berarti \mathbf{w} diberikan oleh

$$\mathbf{w}_{MLE} = \arg \max_{\mathbf{w}} \ln \mathcal{L}(D | \mathbf{w}, \sigma)$$

menjadi

$$\mathbf{w}_{MLE} = \arg \min_{\mathbf{w}} \sum_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2.$$

Selanjutnya turunkan terhadap \mathbf{w} dan membuat sama dengan 0 maka diperoleh

$$\mathbf{w} = (X^T X)^{-1} X^T \mathbf{y}.$$

Berdasarkan MAP, diasumsikan prior Gaussian untuk \mathbf{w} yaitu $w \sim \mathcal{N}(0, \lambda^{-1} I)$

$$P(\mathbf{w}) = \left(\frac{\lambda}{2\pi}\right)^{n/2} \exp\left[-\frac{\lambda}{2} \mathbf{w}^T \mathbf{w}\right]$$

Jadi probabilitas posterior dengan aturan Bayes ditulis sebagai

$$P(\mathbf{w} | D) = \frac{P(\mathbf{w})P(D | \mathbf{w})}{P(D)}$$

Sebagai MLE, logaritma probabilitas posterior dimaksimumkan dan

$$\mathbf{w}_{MAP} = \arg \max_{\mathbf{w}} \ln P(\mathbf{w} | D)$$

menjadi

$$\begin{aligned} \mathbf{w}_{MAP} &= \arg \min_{\mathbf{w}} \sum_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2 \\ &\quad + \frac{\lambda}{2} \mathbf{w}^T \mathbf{w}. \end{aligned}$$

Jadi dengan estimasi MAP dapat sebagai suatu regresi dengan regularisasi. Dugaan MAP pada kasus ini adalah

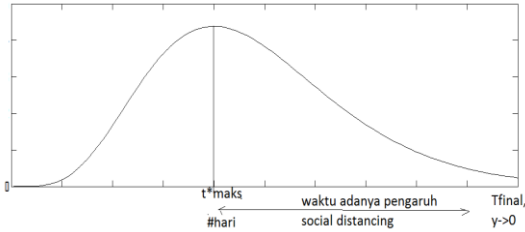
$$\mathbf{w}_{MAP} = (\lambda I + X^T X)^{-1} X \mathbf{y}.$$

Model Eksponensial dan Gaussian

Terdapat beberapa model eksponensial yang digunakan tetapi dapat dipilih bentuk umum yang menunjukkan bahwa total kasus COVID-19 dapat diperoleh, yaitu

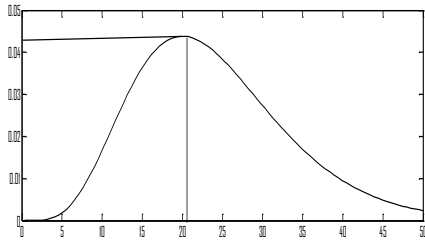
$$y(t) = a_1 + a_2 \exp(a_3 t) \quad (5)$$

Jika model eksponensial digunakan, maka pada waktu tak hingga total kasus COVID-19 juga akan positif tak hingga dimana hal itu tidak mungkin karena upaya pemerintah dalam gerakan *social distancing* tidak memberi dampak sama sekali. Untuk itu model perlu diperbaiki. Diharapkan bahwa *social distancing* atau sekarang lebih disebut dengan *physical distancing* mempunyai pengaruh pada penurunan kasus sebagaimana ditunjukkan pada Gambar 5 dimana model menyerupai Gaussian digunakan.



GAMBAR 5. Ilustrasi model fungsi banyaknya kasus positif COVID-19 di Indonesia sebagai fungsi waktu.

Hingga saat ini t^* maks pada Gambar 5 belum diketahui yaitu waktu dimana total kasus kumulatif di Indonesia mengalami penurunan. Tetapi diharapkan model Gambar 5 memberikan ilustrasi yang berlaku untuk semua kasus COVID-19 yang berhasil diturunkan pandemiknya. Misalkan sekarang bahwa untuk data tak berdimensi, Gambar 5 diperoleh dengan skala pada sumbu horisontal dan vertikal yang ditunjukkan pada Gambar 6.



GAMBAR 6. Ilustrasi model fungsi COVID-19 yang akan dicocokkan skalanya dengan waktu yang ada.

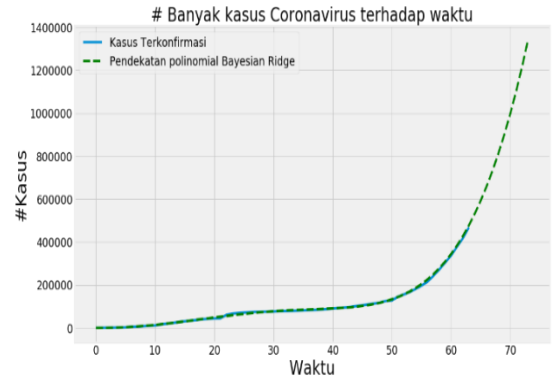
Perhatikan Gambar 5 dan Gambar 6. Akan diduga dimanakah fungsi eksponensial berada pada Gambar 5 jika fungsi eksponensial diterapkan dan menempati pada Gambar 6. Hal inilah yang dilakukan dalam metode ini. Pada model ini data yang digunakan adalah data pada 6 Maret-27 Maret 2020 dan 6 Maret-31 Maret 2020.

HASIL DAN PEMBAHASAN

Hasil SVM dan Ridge Bayesian untuk banyaknya total kasus di dunia

Berdasarkan formulasi (1)-(2) dan bantuan Python, dapat dilakukan pendekatan dan prediksi banyaknya total kasus di dunia pada 22 Januari 2020 hingga 16 Maret 2020. Kemudian dilakukan dugaan untuk 10 hari selanjutnya. Kedua metode SVM dan Bayesian Ridge memberikan prediksi yang

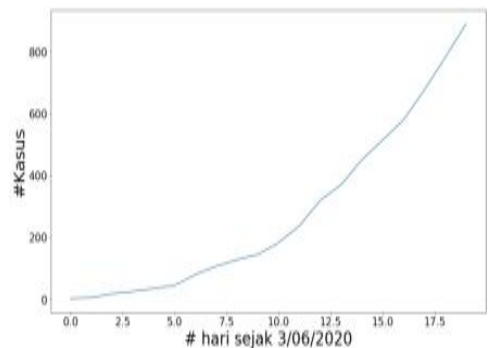
menyerupai dimana semua menunjukkan peningkatan secara eksponensial. Akan tetapi MLE dan MAP untuk Bayesian Ridge lebih kecil dimana pada Gambar 7 terlihat bahwa metode Bayesian Ridge lebih baik.



GAMBAR 7. Laju peningkatan kasus yang dikonfirmasi di dunia yang dicatat WHO sejak 22 Januari 2020 dibandingkan regresi Bayesian Ridge; banyaknya kasus terkonfirmasi (sumbu vertikal), banyaknya hari (sumbu horizontal).

Hasil SVM dan Ridge Bayesian untuk banyaknya total kasus di Indonesia

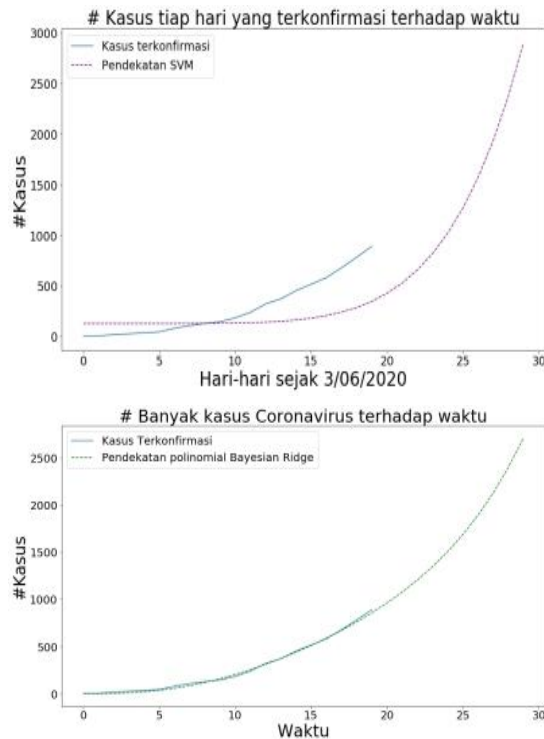
Metode SVM dan Bayesian Ridge telah diujicobakan untuk memodelkan data *time series* untuk sebagian besar negara-negara yang terkonfirmasi adanya paparan virus COVID-19. Kedua metode ini kemudian diterapkan untuk data total kasus positif COVID-19 di Indonesia dengan data dari 22 Januari 2020 hingga 16 maret 2020 yang ditunjukkan pada Gambar 8.



Gambar 8. Banyaknya kasus (sumbu vertikal), banyaknya hari (sumbu horizontal) di Indonesia yang dicatat CNN /dilaporkan Metro TV sejak 6 maret 2020-26 Maret 2020.

Secara sama, maka data dari Indonesia sejak 6 maret 2020 hingga 26-27 maret 2020 dianalisa. Hasil pendekatan Bayesian Ridge memberikan simulasi yang lebih mendekati

data pada Gambar 9 dibandingkan pendekatan SVM. Oleh karena pada kasus dunia dan Indonesia, metode Bayesian Ridge lebih baik, maka data pasien yang meninggal didekati juga menggunakan Bayesian Ridge yang ditunjukkan pada Gambar 10. Akan tetapi dua metode menunjukkan masih adanya peningkatan 10 hari ke depan setelah 26 maret 2020 dengan pendekatan kedua metode.



Gambar 9 (atas). Banyaknya kasus yang dikonfirmasi (sumbu vertikal), banyaknya hari (sumbu horizontal) dan selanjutnya dibandingkan dengan SVM di Indonesia yang dicatat CNN / dilaporkan Metro TV sejak 6 maret 2020 hingga 26 Maret 2020 .

(bawah). Pendekatan kasus yang dikonfirmasi (sumbu vertikal), banyaknya hari (sumbu horizontal) dengan Bayesian Ridge dibandingkan dengan data di Indonesia yang dicatat CNN sejak 6 maret 2020 / dilaporkan Metro TV .

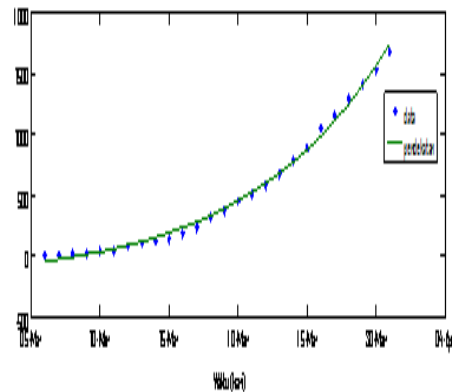
Hasil Model Eksponensial

Beberapa model pertumbuhan COVID-19 di Indonesia telah dilakukan oleh beberapa peneliti di Indonesia maupun komunitas yang terkait dimana laju pertumbuhan telah mengikuti model eksponensial yang beredar di internet. Salah satu model yang bisa digunakan fungsi eksponensial pada persamaan (3). Setelah melakukan *fitting* terhadap data yang ada diperoleh pula hasil yang cocok

dengan data dengan error fungsi 4.38 % yang ditunjukkan pada Gambar 11. Pendekatan ditunjukkan pada data dengan data dari 6 Maret-31 Maret 2020. Akan tetapi dengan model ini, pengguna masih memerlukan waktu selanjutnya beserta prediksi mendatang bagaimana pengaruh *social distancing* terhadap penurunan total kasus COVID-19 di Indonesia. Untuk itulah model eksponensial pada Gambar 10 tidak lagi relevan sekalipun model cukup bagus berdasarkan data yang ada padahal data pada mendatang belum ada. Prediksi demikian dibutuhkan dalam melakukan antisipasi serta menyusun berbagai kebijakan yang dilakukan pemerintah. Data menunjukkan bahwa laju penambahan kasus perhari masih meningkat demikian pula laju total kasus sebagaimana ditunjukkan pada berita metro TV bahwa total kasus 1790 pada 2 April 2020 .



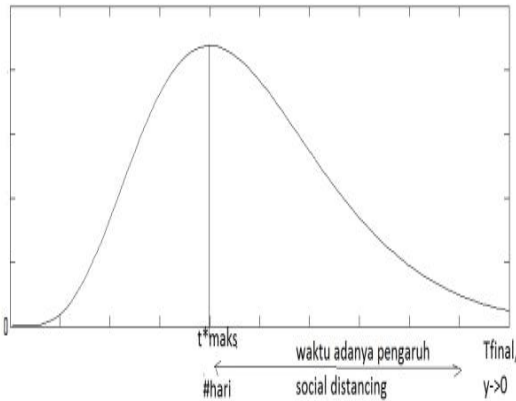
Gambar 10. Pendekatan pasien COVID-19 yang meninggal di Indonesia sejak 6 Maret 2020-27 Maret 2020. (sumbu vertikal), banyaknya hari (sumbu horizontal) di Indonesia yang dicatat CNN dan berita metro TV sejak 6 Maret 2020-27 Maret 2020 dengan Bayesian Ridge dibandingkan dengan data.



Gambar 11. Model pendekatan eksponensial untuk data total konfirmasi data per hari di Indonesia dari 6 maret 2020-31 maret 2020.

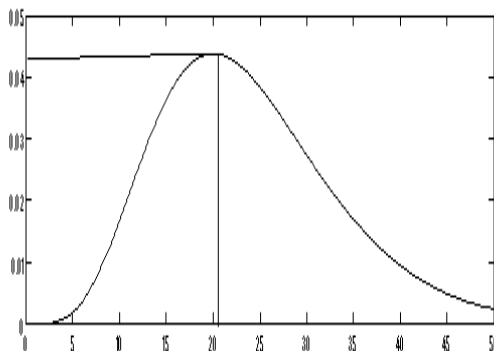
Hasil model Gaussian beserta prediksi waktu

Sebagaimana diketahui bahwa *social distancing* (SD) atau sekarang lebih disebut dengan *physical distancing* mempunyai pengaruh pada penuruan kasus. Diharapkan bahwa laju pertumbuhan total kasus dengan adanya SD mengikuti pola yang ditunjukkan pada Gambar 12 dimana model menyerupai Gaussian digunakan.



Gambar 12 Ilustrasi model fungsi COVID 19 di Indonesia sebagai fungsi waktu.

Hingga saat ini t^*maks belum diketahui yaitu waktu dimana total kasus kumulatif di Indonesia mengalami penurunan. Tetapi diharapkan model Gambar 12 memberikan ilustrasi yang berlaku untuk semua kasus COVID-19 yang berhasil diturunkan pandemiknya. Misalkan sekarang bahwa untuk data tak berdimensi, Gambar 12 diperoleh dengan skala pada sumbu horisontal dan vertikal dengan Gambar 13.



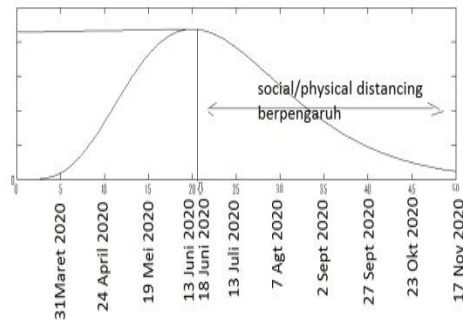
Gambar 13. Ilustrasi model fungsi COVID 19 yang akan dicocokkan skalanya dengan waktu yang ada.

Perhatikan Gambar 11 dan Gambar 13 kemudian ingin diduga dimanakah fungsi

eksponensial berada pada Gambar 11 jika Gambar 11 diskalakan pada Gambar 13. Terdapat 2 hasil prediksi yang diperoleh yang dijelaskan pada berikut ini.

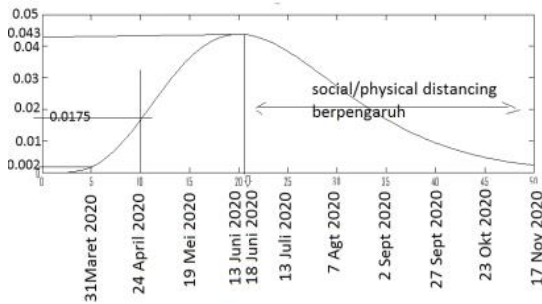
Skenario Terburuk Waktu Pertumbuhan COVID-19 Di Indonesia

Jika dalam 25 hari berlaku data awal Indonesia (sebenarnya digunakan 26 hari, untuk mempermudah penskalaan), pada skala 0-5 pada Gambar 13 maka pada point sumbu horizontal adalah 5, maka waktu terakhir pada Gambar 11 berada pada tanggal 31 Maret 2020. Karena setiap penambahan 5 satuan pada Gambar 11 berarti 25 hari maka pada puncak berlaku sekitar 105 hari. Demikian dilakukan maka diperoleh skenario kasus COVID-19 di Indonesia dengan waktunya ditunjukkan pada Gambar 14.



Gambar 14. Ilustrasi skenario terburuk model fungsi COVID-19 di Indonesia yang akan dicocokkan skalanya dengan waktu yang ada dengan pengaruh *social distancing* sudah berpengaruh .

Berdasarkan model itu, total kasus pada tanggal -tanggal tertentu di atas dapat ditentukan. Karena berdasarkan data diketahui bahwa pada 31 maret 2020 sebanyak 1677 kasus, maka sekarang dapat diduga total kasus positif COVID-19 pada tanggal yang lain. Misalkan pada 31 maret maka sumbu y bernilai 0.002 . Oleh karena itu pada 24 April 2020 akan berada pada tinggi sumbu y yang bernilai 0.0175. Dengan skala tersebut dapat diperkirakan bahwa banyaknya total kasus hingga 24 April 2020 sekitar 14675 kasus yang dihasilkan pada Gambar 15.

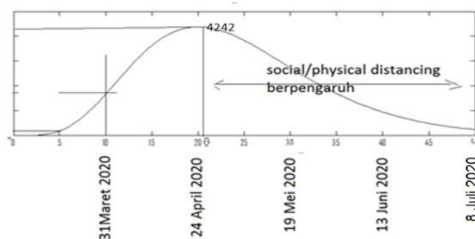


Gambar 15. Ilustrasi skenario model fungsi COVID 19 di Indonesia yang akan dicocokkan skalanya dengan waktu yang ada dengan pengaruh social distancing sudah berpengaruh dengan skala banyaknya total kasus.

Nampaknya skenario terburuk tidak terjadi karena dalam 25 hari selanjutnya menjadi 14675 dianggap terlalu besar dari pengamatan data yang ada. Untuk itu tanggal digeser pada sumbu horizontal dengan model berikut.

Skenario Kedua yang Lebih Realistis

Data dari 6 maret 2020-hingga 31 Maret 2020 berada pada skala 10 satuan pada sumbu horizontal. Dengan skenario ini maka diperoleh urutan tanggal dan puncak terjadi pada 24 April 2020 dan ditunjukkan pada Gambar 16.



Gambar 16. Ilustrasi skenario realistis dengan total kasus pada puncak di 24 April 2020 dengan adanya pengaruh *social distancing*

Menggunakan skenario lebih realistis ini, dapat diduga banyaknya total kasus hingga pada 24 April 2020 sebesar 4242 dimana hasil ini lebih cukup realistis setelah 25 hari dari 31 Maret 2020 dan ini merupakan waktu dimana terjadi puncak total kasus sehingga *social distancing* baru mulai memberikan dampaknya bagi pertumbuhan total kasus COVID-19 di Indonesia. Dibandingkan dengan penulisan dalam beberapa peneliti dalam media massa (Ardila, n.d.) maka hasil ini sangat jauh karena pada April 2020 diprediksi oleh peneliti lain dapat

sekitar 500 ribu. Sedangkan menurut peneliti dari ITB menyatakan bahwa puncak terjadi pada akhir Maret 2020 dan berakhir pada pertengahan April dengan kasus harian sekitar 600 (Adi, n.d.). Hingga pada 2 April 2020, kasus harian adalah 113 sehingga dapat dimungkinkan bahwa prediksi pada pertengahan April 2020 masih sangat jauh dari 600. Akan tetapi hingga artikel ini ditulis, telah diketahui bahwa data total kasus positif COVID-19 di Indonesia adalah 8211. Dengan membandingkan hasil prediksi dan data yang diperoleh, diketahui bahwa pendekatan yang dilakukan belum mendekati dengan data sehingga masih perlu diperbaiki.

KESIMPULAN

Total kasus di dunia menggunakan data sejak 22 Januari 2020 hingga 16 Maret 2020, kemudian dilakukan pendekatan untuk mendapatkan total kasus 10 hari selanjutnya dengan menggunakan SVM dan Bayesian Ridge. Kedua metode menunjukkan bahwa terjadi peningkatan pada 10 hari selanjutnya dimana Bayesian Ridge memberi prediksi lebih baik. Selanjutnya kedua metode digunakan untuk memprediksi data total kasus positif COVID-19 di Indonesia dengan data dari tanggal 6 maret 2020-27 Maret 2020 dimana prediksi Bayesian Ridge terbaik dengan MLE dan MAP yang lebih kecil. Untuk melibatkan pengaruh *social distancing*, digunakan model eksponen untuk mencari bentuk awal Gaussian dan kemudian digunakan untuk melakukan prediksi pada waktu tertentu dengan data pada tanggal 6 maret 2020-31 Maret 2020. Model Gaussian *like* yang dihasilkan dengan skala yang sudah ada dan ditransformasikan sesuai dengan data. Model ini diusulkan dengan 2 skenario dimana skenario terburuk memberikan hasil bahwa *social distancing* mulai memberi dampak pada mulai sekitar 18 Juni 2020. Akan tetapi total kasus yang diprediksi pada 24 April 2020 sekitar 14675 kasus. Jika diperhatikan dari data yang sudah ada dalam 25 hari pada 6 maret-31 maret 2020, total kasus adalah 1677 maka prediksi skenario terburuk tidak dapat diterima. Untuk itu maka diusulkan skenario kedua dimana puncak total kasus COVID-19 di Indonesia terjadi pada 24 April 2020 dengan total kasus 4242. Akan tetapi dibandingkan data total kasus pada 24 april 2020, maka

prediksi yang diperoleh masih perlu diperbaiki.

UCAPAN TERIMA KASIH

Penelitian ini didanai oleh pusat riset penelitian UKSW tahun anggaran 2019/2020 dalam rangka kepedulian terhadap kasus COVID-19 di dunia dan di Indonesia.

DAFTAR PUSTAKA

- Adi, P. (n.d.). *Peneliti ITB Buat Pemodelan Prediksi Puncak Penyebaran COVID-19 di Indonesia*. <https://www.itb.ac.id/news/read/57444/home/peneliti-itb-buat-pemodelan-prediksi-puncak-penyebaran-covid-19-di-indonesia>.
- Ardila, S. (n.d.). *Bappenas, UI modeling shows grim projection of COVID-19 spread in Indonesia*. <https://www.thejakartapost.com/news/2020/03/31/bappenas-ui-modeling-shows-grim-projection-of-covid-19-spread-in-indonesia.html>.
- Demidova, L., E. Nikulchev & Y. Sokolova. 2016. Big Data Classification Using the SVM Classifiers with the Modified Particle Swarm Optimization and the SVM Ensembles. *International Journal of Advanced Computer Science and Applications*, **7(5)**: 294–312. <https://doi.org/10.14569/ijacsa.2016.070541>.
- Hamid, A.R.A.H. 2020. Social responsibility of medical journal: a concern for COVID-19 pandemic. *Medical Journal of Indonesia*, **29(1)**:1–3. <https://doi.org/10.13181/mji.ed.204629>.
- Kowalczyk, A. 2017. *Support Vector Machines Succinctly*. www.syncfusion.com.
- Lamidi, A. 2018. *No Title*. <https://towardsdatascience.com/breast-cancer-classification-using-support-vector-machine-svm-a510907d4878>.
- Mckinney, W. 2018. *Python for Data Analysis*, 2nd edition. O Reilly Media, Inc.
- Min-Wei, H., C. Chih-Wen, L. Wei-Chao, K. Shih-Wen & T. Chih-Fong. 2018. SVM and SVM Ensembles in Breast Cancer Prediction. *Plos One*, **12**. <https://doi.org/10.1371/journal.pone.0161501>.
- Ortiz-Ospina, M.R.H.R.E. 2020. *Coronavirus Disease (COVID-19)–Statistics and Research*. <https://ourworldindata.org/coronavirus>.
- Pambuccian, S. E. 2020. The COVID-19 pandemic: Implications for the cytology laboratory. *Journal of the American Society of Cytopathology*. <https://doi.org/https://doi.org/10.1016/j.jasc.2020.03.001>.
- Permai, S.D. & H. Tanty. 2018. Linear regression model using bayesian approach for energy performance of residential building. *Procedia Computer Science*, **135**: 671–677. <https://doi.org/10.1016/j.procs.2018.08.219>.
- Prem, K., Y. Liu, T. Russell, A.J. Kucharski, R.M. Eggo, N. Davies, C. Group, for the M.M. of I.D.C.-19 W., M. Jit & P. Klepac. 2020. The effect of control strategies that reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China. *MedRxiv*, **2667(20)**: 2020.03.09.20033050. <https://doi.org/10.1101/2020.03.09.20033050>.
- Qian, H. 2018. Big data Bayesian linear regression and variable selection by normal-inverse-gamma summation. *Bayesian Analysis*, **13(4)**: 1007–1031. <https://doi.org/10.1214/17-BA1083>.
- Web1. <https://github.com/CSSEGISandData/COVID-19>.