BAB II

LANDASAN TEORI

2.1 Corona Disease 2019 (COVID – 19)

Coronavirus Disease 2019 (COVID-19) adalah jenis virus baru pertama kali ditemukan di kota Wuhan, China pada 31 Desember 2019 (WHO, 2020). Presiden Indonesia melaporkan kasus pertama yang terjadi di Indonesia pada tanggal 2 Maret 2020 (Ihsanuddin, 2020). Dokter Merry Dame Cristy (Pane, 2020) mengatakan, penyakit ini bisa menyebabkan infeksi pernapasan berat, seperti infeksi paru-paru(*pneumonia)*.

Virus ini menular melalui percikan dahak (*droplet*) dari saluran pernapasan, penularan dapat terjadi saat seorang menghirup *droplet* dari penderita atau apabila *droplet* terkena mata maka dapat beresiko tertular pernyakit ini (Limbong, 2020). Gejala awal yang dirasakan jika terinfeksi virus ini seperti terkena gejala flu, yaitu demam, pilek, batuk kering, sakit tenggorokan dan sakit kepala (Pane, 2020).

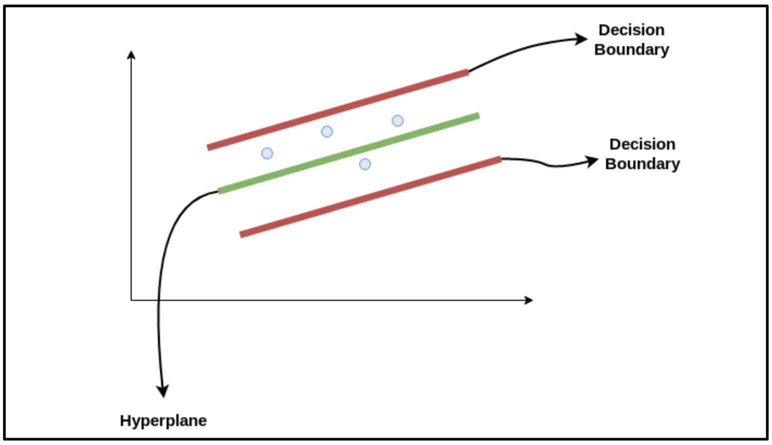
Menurut Nur Rohim Yunus dan Annisa Rezki (Yunus & Rezki, 2020), perkembangan penularan COVID-19 cukup signifikan. Data yang dimiliki Satuan Tugas Penanganan COVID-19 (Satgas COVID-19, 2021), tercatat hingga tanggal 31 Agustus 2020 ada 174.554 kasus positif di Indonesia. Dalam usaha untuk mencegah penularan kasus virus corona, pemerintah membuat kebijakan untuk menanganinya (Yunus & Rezki, 2020). Usaha pemerintah dalam menetapkan kebijakan seperti Pembatasan Sosial Berskala Besar (PSBB), berdiam di rumah, menjaga kebersihan, bekerja dan belajar dari rumah untuk menangani penyebran virus corona (Tuwu, 2020). Protokol kesehatan juga sudah ditetapkan oleh pemerintah sebagai salah satu cara mengurangi penularan seperti memakai masker, mencuci tangan, dan menjaga jarak (Mardhia et al., 2020).

2.1 Regresi

Regresi adalah proses identifikasi relasi dan pengaruhnya pada nilai-nilai objek (Suyanto, 2018). Regresi bertujuan untuk menemukan fungsi yang memodelkan data dengan meminimalkan selisih antar nilai prediksi (Suyanto, 2018). Pada umumnya regresi digunakan untuk prediksi (*prediction*) dan peramalan (*forecasting*). Prediksi digunakan untuk memperkirakan nilai-nilai data yang bertipe apa dan kapan saja (masa lalu, masa depan, dan sekarang) (Suyanto, 2018).

2.2 Support Vector Regression

Support Vector Regression (SVR) adalah versi lain dari *Support Vector Machine* (SVM) yang digunakan untuk memecahkan masalah regresi. Tujuan SVR adalah untuk menemukan fungsi seperti *hyperplane* (garis pemisah) dalam bentuk fungsi regresi yang akan cocok dengan semua *input* dengan sebuah *error* dan membuatnya setipis mungkin (Smola & Scholkopf, 2002).



Gambar 2.1 Hyperplane Support Vector Regression
(Sethi, 2020)

Fungsi persamaan *hyperplane* pada Gambar 2.1. Diformulasikan sebagai berikut (Sethi, 2020):

$$f(x) = w^T \varphi(x) + b \qquad \ldots(2.1)$$

Dengan:

w^T = vektor bobot

$\varphi(x)$ = fungi yang memetakan nilai x dalam suatu dimensi

b = bias

Menurut Caraka et al. (2017), agar mendapatkan generalisasi yang baik terhadap fungsi *f(x)*, dapat dilakukan dengan meminimalkan *norm* dari *w* dengan penyelesaian optimasi dengan formulai:

$$\min \{\frac{1}{2}\|w\|^2\} \qquad \ldots(2.2)$$

Dengan syarat:

$y_i - w^T\varphi(x_i) - b \leq \varepsilon$, $i = 1, 2, \ldots, l$

$w^T\varphi(x_i) - y_i + b \leq \varepsilon$, $i = 1, 2, \ldots, l$

Diasumsikan bahwa terdapat fungsi *f(x)* yang dapat mengaproksimasi semua titik (x_i, y_i), dengan presisi ε. Pada kasus ini diasumsikan bahwa semua titik dalam rentang *f(x)* ± ε (*feasible*). Dalam hal ketidaklayakan (*infeasible*), ada beberapa titik yang kemungkinan keluar dari rentang *f(x)* ± ε, oleh karena itu dapat ditambahkan variable ζ, ζ^* untuk mengatasi masalah ketidaklayakan dalam problem optimasi. Masalah optimasi di atas dapat diformulasikan sebagai berikut:

$$\min \{\frac{1}{2}\|w\|^2 + C\sum_{i}^{l}(\zeta_i + \zeta_i^*)\}$$

…(2.3)

Dengan syarat:

$y_i - w^T\varphi(x_i) - b - \zeta_i \leq \varepsilon$, i = 1, 2, …, l

$w^T\varphi(x_i) - y_i + b - \zeta_i^* \leq \varepsilon$, i = 1, 2, …, l

$\zeta, \zeta_i^* \geq 0$

Menurut (Smola & Scholkopf, 2002), solusi optimal untuk persamaan (2.2) dapat dilakukan dengan bentuk formulasi sebagai berikut:

$$max \begin{cases} -\dfrac{1}{2}\displaystyle\sum_{i.j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ -\varepsilon\displaystyle\sum_{t=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{t=1}^{l} y_t(\alpha_i - \alpha_i^*) \end{cases} \quad \ldots(2.4)$$

Dengan syarat:

$$\sum_{t=1}^{l}(a_t - a_t^*) = 0 \; dan \; \alpha_t, a_t^* \in [0, C]$$

Dengan begitu fungsi regresi *f(x)* dapat dituliskan sebagai berikut:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - a_i^*)K(x_i, x) + b \quad \ldots(2.5)$$

Dimana α_i dan a_i^* adalah variabel *Lagrange* dan $(\alpha_i - a_i^*)$ merupakan nilai beta, $K(x_i, x)$ adalah fungsi kernel dan b adalah *bias*. Menurut (Smola & Scholkopf, 2002) solusi optimal untuk *bias* dapat dihitung menggunakan *Karush-Kuhn-Tucker* (KKT), perhitungan KTT adalah sebagai berikut:

$$\alpha_i(\varepsilon + t_i - y_i + w^T\varphi(x) + b) = 0 \quad \ldots(2.6)$$

$$\alpha_i(\varepsilon + t_i - y_i + w^T\varphi(x) - b) = 0 \quad \ldots(2.7)$$

$$(C - a_i)t_i = 0 \quad \ldots(2.8)$$

$$(C - a_i^*)t_i = 0 \quad \ldots(2.9)$$

Sehingga didapatkan:

$$b = y_i - w^T\varphi(x) - \varepsilon \text{ untuk } 0 < a_i < C \qquad \ldots(2.10)$$

$$b = y_i - w^T\varphi(x) + \varepsilon \text{ untuk } 0 < a_i^* < C \qquad \ldots(2.11)$$

2.3 Kernel

Kernel yang digunakan pada *Support Vector Regression* (SVR) merupakan fungsi yang dapat memetakan data ke ruang fitur yang memiliki dimensi lebih tinggi agar data dapat diproses dengan lebih terstruktur (Bonita et al., 2018). Pemetaan data dapat ke dimensi yang lebih tinggi melalui φ (Amanda et al., 2014). Pada penelitian ini akan menggunakan kernel *Linear*, *Polynomial*, *dan Radial Basis Function* (RBF), berikut formulasi untuk ketiga kernel tersebut:

1. Linear Kernel

$$K(x, y) = x\,.\,y \qquad \ldots(2.12)$$

Dengan *x* dan *y* merupakan data yang akan diklasifikasikan

2. Polynomial Kernel

$$K(x, y) = (1 + x\,.\,y)^d \qquad \ldots(2.13)$$

Dengan x dan y merupakan data yang akan diklasifikasikan, dan *d* merupakan derajat polynomial.

3. Radial Basis Function

$$K(x, y) = exp\,(-\frac{\left||x - x'|\right|^2}{2\sigma^2}) \qquad \ldots(2.14)$$

Dengan σ merupakan varians dan hyperparameter, $x - x'$ merupakan jarak antara dua poin *x*.

2.4 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) adalah salah satu metode yang digunakan untuk menghitung *error* yang dihasilkan prediksi (*predict)*. MAPE dihitung menggunakan kesalahan absolut lalu dibagi dengan nilai yang diamati (Nabillah & Ranggadara, 2020). MAPE menunjukkan seberapa besar kesalahan dalam memprediksi dibandingkan dengan nilai sebenarnya (Khair et al., 2017).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| x\ 100\% \qquad \ldots(2.15)$$

Dimana *n* merupakan jumlah data, *A* adalah merupakan hasil aktual, dan *F* merupakan nilai prediksi.

MAPE memiliki kriteria nilai yang berguna untuk mengukur performa yang dihasilkan. Kriteria keakuratan MAPE ditunjukkan pada Tabel 1.

Tabel 2.1 Kriteria Keakuratan MAPE

Nilai MAPE	Kriteria
< 10%	Sangat Baik
10% - 20%	Baik
20% - 50%	Cukup
>50%	Tidak Akurat

2.5 Mean Absolute Error

Mean Absolute Error (MAE) adalah salah satu teknik yang digunakan untuk menghitung keakuratan model prediksi (Suryanto, 2019). MAE merupakan nilai rata-rata dari *error* yang diabsolutkan, dimana *error* merupakan selisih dari nilai sebenarnya dengan nilai hasil prediksi (Irfan et al., 2014).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - R_i| \quad \dots(2.16)$$

Dengan:

n = jumlah data

P = nilai hasil peramalan

R = nilai sebenarnya

Berdasarkan rumus diatas, MAE melakukan perhitungan rata-rata *error* dengan memberikan bobot yang sama pada seluruh data yang ada. Untuk melakukan evaluasi *prediction*, MAE lebih intuitif dalam memberikan rata-rata *error* dari keseluruhan data (Suryanto, 2019).

2.6 Mean Squared Error

` *Mean Squared Error* (MSE) merupakan metode evaluasi lainnya. Metode MSE melakukan perhitungan evaluasi dengan melakukan pengurangan kuadrat dari nilai sebenarnya dengan nilai prediksi (Azmi et al., 2020). Formulasi MSE adalah sebagai berikut:

$$MSE = \frac{\sum_{i=1}^{n}(X_i - F_i)}{n} \quad \dots(2.17)$$

Dimana:

n = jumlah data

Xi = nilai sebenarnya

Fi = nilai prediksi