

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek dalam penelitian kali ini difokuskan untuk meneliti data *time series* mengenai kasus positif aktif harian Covid-19 dari tempat tinggal dari pengunjung Pusat Kebugaran X yang meliputi provinsi DKI Jakarta, Banten dan Jawa Barat dari bulan Maret 2020 hingga hari ini secara *real time*. Hal pertama yang akan dilakukan adalah menganalisis prediksi total kasus aktif harian Covid-19 di beberapa provinsi tempat tinggal para pengunjung Pusat Kebugaran X yaitu provinsi DKI Jakarta, Banten dan Jawa Barat selama 1 bulan ke depan setelah hari ini secara *real time*, setelah itu dari hasil prediksi tersebut akan dilakukan *clustering* berdasarkan tanggal-tanggal, mulai dari tanggal yang memiliki tingkatan total kasus aktif harian yang cukup tinggi hingga tanggal yang memiliki tingkatan total kasus aktif harian yang cukup rendah. Dari hasil *clustering* tersebut akan dibentuk sebuah visualisasi *dashboard* yang menentukan apakah para pengunjung dari Pusat Kebugaran X boleh mengunjungi Pusat Kebugaran X atau tidak dalam suatu periode waktu tertentu. Penerapan saran kebijakan yang dilakukan Pusat Kebugaran X ini diharapkan dapat membuat para pengunjung serta staf Pusat Kebugaran X tetap merasa aman, nyaman dan tidak cemas terhadap penyebaran Covid-19 meskipun mereka sedang beraktivitas di Pusat Kebugaran X.

Dataset yang digunakan dalam penelitian kali ini merupakan data dalam bentuk *time series* berupa file *excel*. Data yang digunakan merupakan data kasus Covid-19 pada beberapa provinsi yang menjadi tempat tinggal daripada para pengunjung dari Pusat Kebugaran X yaitu provinsi DKI Jakarta, Banten dan Jawa Barat dari bulan Maret 2020 hingga hari ini secara *real time*. Untuk data Covid-19 dari provinsi DKI Jakarta terdiri dari 9 kolom yang terdiri sebagai berikut:

- Tanggal : Tanggal terjadinya kasus Covid-19.
- Meninggal : Jumlah penderita kasus Covid-19 yang meninggal.

- Sembuh : Jumlah penderita kasus Covid-19 yang sembuh.
- Self Isolation : Jumlah penderita kasus Covid-19 yang sedang melakukan isolasi mandiri.
- Masih Perawatan : Jumlah penderita kasus Covid-19 yang sedang masih dalam perawatan.
- Total Pasien : Jumlah total pasien yang menderita kasus Covid-19.
- Positif Harian : Jumlah penambahan kasus positif Covid-19 per hari.
- Positif Aktif : Jumlah kasus positif Covid-19 yang sedang aktif.
- Sembuh Harian : Jumlah penambahan kasus sembuh Covid-19 per hari.

Tabel 3.1 berikut merupakan sepuluh baris pertama mengenai data kasus harian Covid-19 dari provinsi DKI Jakarta yang sudah didapatkan.

Tabel 3.1 Data Kasus Harian Covid-19 DKI Jakarta

Tanggal	Meninggal	Sembuh	Self Isolation	Masih Perawatan	Total Pasien	Positif Harian	Positif Aktif	Sembuh Harian
01/03/2020	0	0	0	0	0	0	0	0
02/03/2020	0	0	0	0	0	0	0	0
03/03/2020	1	0	0	2	3	3	2	0
04/03/2020	1	0	0	2	3	0	2	0
05/03/2020	3	0	0	4	7	4	4	0
06/03/2020	3	0	0	4	7	0	4	0
07/03/2020	3	0	0	4	7	0	4	0
08/03/2020	3	0	0	4	7	0	4	0
09/03/2020	3	0	0	31	34	27	31	0
10/03/2020	3	0	0	31	34	0	31	0

Di sisi lain untuk data Covid-19 dari provinsi Banten dan Jawa Barat hanya terdiri dari 4 kolom. Kolom-kolom tersebut adalah sebagai berikut:

- Tanggal : Tanggal terjadinya kasus Covid-19.
- Positif Aktif : Jumlah kasus positif Covid-19 yang sedang aktif.
- Meninggal : Jumlah penderita kasus Covid-19 yang meninggal.
- Sembuh : Jumlah penderita kasus Covid-19 yang sembuh.

Tabel 3.2 dan Tabel 3.3 berikut merupakan 10 baris pertama mengenai data kasus harian Covid-19 dari provinsi DKI Jakarta yang sudah didapatkan:

Tabel 3.2 Data Kasus Harian Covid-19 Banten

Tanggal	Positif Aktif	Sembuh	Meninggal
21/03/2020	40	1	2
22/03/2020	43	1	3
23/03/2020	52	1	3
24/03/2020	60	1	4
25/03/2020	62	1	4
26/03/2020	62	1	4
27/03/2020	79	1	4
28/03/2020	98	1	4
29/03/2020	101	1	4
30/03/2020	123	1	4

Tabel 3.3 Data Kasus Harian Covid-19 Jawa Barat

Tanggal	Positif Aktif	Sembuh	Meninggal
21/03/2020	47	1	7
22/03/2020	45	5	9
23/03/2020	45	5	9
24/03/2020	45	5	10
25/03/2020	58	5	10
26/03/2020	62	5	11
27/03/2020	79	5	14
28/03/2020	96	6	17
29/03/2020	124	6	19
30/03/2020	152	8	20

3.2 Metode Penelitian

Pada penelitian ini akan menggunakan metode *Knowledge Discovery in Database* (KDD) sebagai langkah-langkah atau kerangka yang digunakan dalam pengerjaan penelitian ini dengan menggunakan bantuan *tools Rapid Miner*. *Tools Rapid Miner* dipilih karena memiliki waktu pemrosesan data yang lebih cepat dan memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan WEKA. Untuk lebih jelasnya dapat dilihat pada Tabel 3.4 di bawah.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tabel 3.4 Perbandingan Rapid Miner dan WEKA

Faktor Pembeda	Rapid Miner	WEKA
Kecepatan Pemrosesan Data	Lebih cepat 2,6% dibandingkan dengan WEKA	Lebih lambat 2,6% dibandingkan dengan <i>Rapid Miner</i> [41].
Tingkat Akurasi Metode Klasifikasi	Lebih tinggi dibandingkan WEKA (> 90%)	Lebih rendah dibandingkan <i>Rapid Miner</i> (<90%) [42]

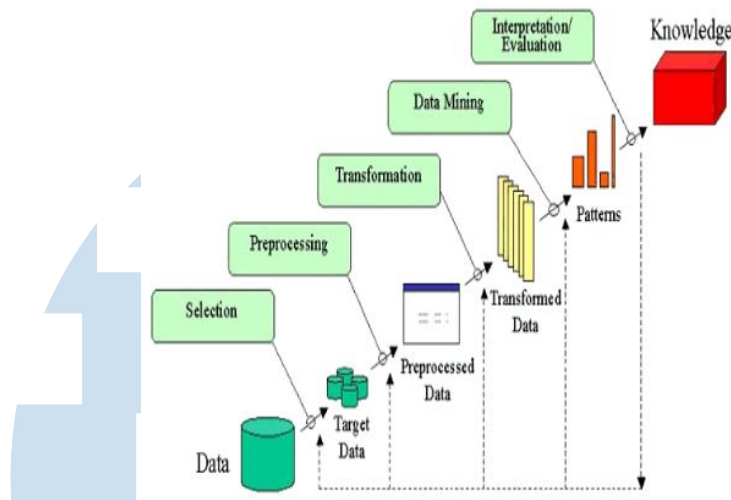
Di sisi lain, metode KDD ini dipilih karena metode ini biasa digunakan oleh para peneliti dan mencari informasi tersembunyi dalam suatu data. Sesuai dengan penelitian ini yang dilakukan oleh peneliti dan bukan untuk kepentingan bisnis dan memecahkan masalah bisnis melainkan untuk mengetahui informasi tersembunyi dalam sebuah data yaitu mengenai analisis prediksi dan *clustering*. Untuk lebih lengkapnya dapat dilihat pada Tabel 3.5 berikut ini.

Tabel 3.5 Perbandingan Metode KDD dengan CRIPS-DM

Faktor Pembeda	<i>Knowledge Discovery in Database</i>	CRISP-DM
Pengguna	Biasa digunakan oleh para data <i>mining</i> expert dan para peneliti	Biasa digunakan oleh perusahaan
Tujuan	Untuk mencari informasi tersembunyi/ pola di dalam data	Untuk memberikan solusi terhadap masalah bisnis
Jumlah tahapan	5 Tahap	6 Tahap

Sumber: [43]

Secara umum, tahapan-tahapan yang akan dilakukan dalam metode KDD dapat dilihat pada Gambar 3.1 berikut.



Gambar 3.1 Tahapan dalam Metode KDD

Sumber: [44]

Berikut merupakan penjelasan dari tahapan *Knowledge Discovery in Database* seperti gambar di atas yang akan digunakan dalam penelitian ini:

3.2.1 Data Selection

Dalam sub-metode ini merupakan tahapan dalam pemilihan data yang akan digunakan untuk proses peramalan. Data mengenai kasus Covid-19 pada beberapa provinsi tempat tinggal dari para pengunjung Pusat Kebugaran X semuanya sudah didapatkan. Semua data didapatkan dalam bentuk *file excel*. Dari masing-masing *file* tersebut terdapat beberapa kolom. Namun tidak semua kolom akan digunakan, hanya kolom 'Tanggal' dan 'Positif Aktif' saja yang akan digunakan untuk prediksi dengan metode *Holt-Winters*. Sedangkan dalam proses *clustering* akan digunakan kolom tanggal beserta dengan kolom yang berisi hasil prediksi yang sudah didapatkan dari proses peramalan yang dilakukan dengan menggunakan metode *Holt-Winters*.

3.2.2 Preprocessing

Pada sub-metode ini berfokus dalam melakukan proses persiapan data yang sudah dipilih yaitu kolom Tanggal dan Positif Aktif. Proses persiapan data yang dimaksud merupakan proses data *cleaning*. Dari data

pada kolom Tanggal dan Positif Aktif yang memiliki nilai *null* akan dilakukan proses *data cleaning*. Selain itu juga akan dilakukan pengecekan apakah ada duplikasi data atau tidak. Proses *cleaning* pada data akan dilakukan dengan menggunakan *tools Rapid Miner*.

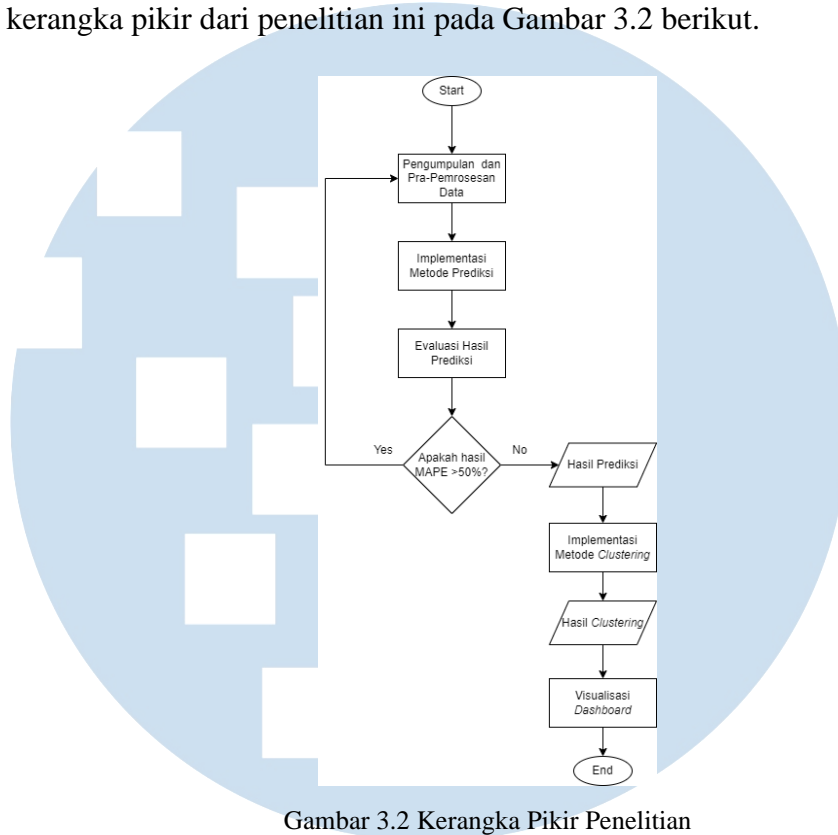
3.2.3 Transformation

Tahap dari sub-metode ini merupakan proses transformasi untuk mengubah bentuk data yang belum siap untuk dilakukan proses data *mining* ke dalam bentuk data yang sudah siap untuk dilakukan proses data *mining*. Data yang sudah dilakukan proses data *cleaning* selanjutnya akan dibentuk ke dalam data *time series*. Di mana kolom *Date* yang akan dijadikan acuan untuk proses transformasi bentuk data ke dalam bentuk data *time series*. Selain itu juga akan dilakukan pengecekan apakah data yang digunakan cenderung fluktuatif atau cenderung konstan. Apabila data yang digunakan cenderung fluktuatif maka metode *Holt-Winters* yang digunakan adalah metode *Holt-Winters* multiplikatif, namun apabila data yang digunakan cenderung konstan maka metode *Holt-Winters* yang digunakan adalah metode *Holt-Winters* aditif.

3.2.4 Data Mining

Pada sub-metode Data *Mining* ini merupakan sub-metode yang paling penting dalam kerangka KDD karena pada tahap ini akan dilakukan pencarian informasi yang tersembunyi di dalam data dengan menggunakan metode tertentu. Pada penelitian kali ini terdapat dua metode yang digunakan yaitu metode *Holt-Winters* untuk prediksi dan metode *K-Means* untuk *clustering*. Metode *Holt-Winters* dilakukan untuk mendapatkan hasil peramalan, kemudian hasil peramalan tersebut akan dijadikan sebagai *input* pada proses *clustering*. Setelah proses *clustering* selesai dilakukan maka *output*-nya akan menjadi *input* dalam proses visualisasi yang akan menjadi *dashboard* yang memberikan saran kebijakan kepada para pengunjung Pusat Kebugaran X apakah mereka boleh mengunjungi Pusat Kebugaran X

dalam periode waktu tertentu. Untuk lebih jelasnya dapat dilihat pada kerangka pikir dari penelitian ini pada Gambar 3.2 berikut.



Gambar 3.2 Kerangka Pikir Penelitian

Pemilihan metode *Holt-Winters* untuk melakukan prediksi karena metode *Holt-Winters* dapat menerima semua jenis data baik yang stasioner maupun tidak stasioner, di mana metode *Holt-Winters* dapat menangani data yang memiliki pola musiman atau mempunyai *trend*. Sesuai dengan data yang digunakan dalam penelitian ini yang memiliki *trend*, tidak stasioner dan cenderung fluktuatif. Untuk lebih jelasnya dapat dilihat pada Tabel 3.6 berikut ini.

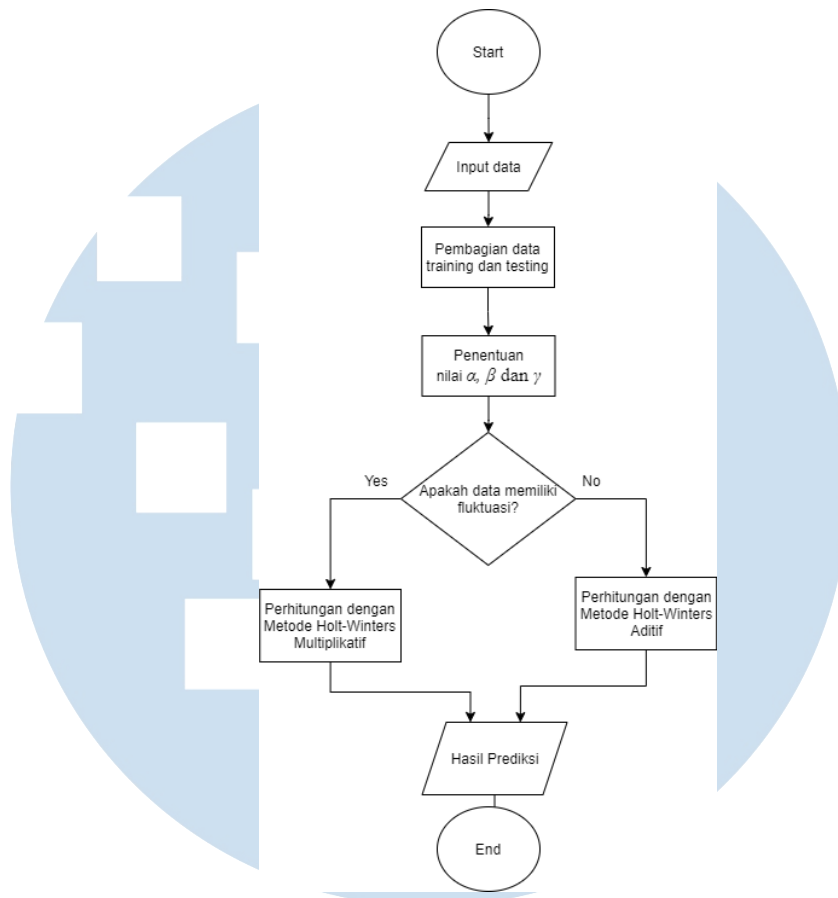
Tabel 3.6 Perbandingan Metode Holt-Winters dengan ARIMA

Faktor Pembeda	Holt-Winters	ARIMA
Pengertian	Metode yang digunakan untuk mengatasi data musiman dan data yang	Metode yang tidak memedulikan variabel independen dalam perhitungan

Faktor Pembeda	Holt-Winters	ARIMA
	memiliki unsur <i>trend</i> [45]	peramalan dan sangat cocok untuk peramalan jangka pendek [46]
Pola Data yang diterima	Data stasioner, data tidak stasioner, data musiman dan data yang memiliki <i>trend</i> [25]	Data yang stasioner [47]

Secara garis besar Gambar 3.3 berikut merupakan langkah-langkah penggunaan metode *Holt-Winters* dalam sub-metode *Data Mining*.





Gambar 3.3 Langkah Metode Holt-Winters

Setelah data sudah dilakukan proses data *cleaning* dan *transformation*, maka langkah pertama dalam proses *data mining* adalah dengan membagi data menjadi data uji dan data latih. Setelah itu akan dilakukan penentuan nilai α , β dan γ . Langkah selanjutnya yaitu menentukan apakah data yang digunakan memiliki fluktuasi atau tidak, apabila data yang digunakan memiliki fluktuasi maka akan dilakukan perhitungan menggunakan metode *Holt-Winters* multiplikatif sedangkan apabila data yang digunakan tidak memiliki fluktuasi maka akan dilakukan perhitungan menggunakan metode *Holt-Winters* aditif. Perhitungan dengan metode *Holt-Winters* akan menghasilkan *output* berupa hasil prediksi yang nantinya akan dievaluasi tingkat akurasinya dan dijadikan sebagai *input* dalam metode *K-Means*.

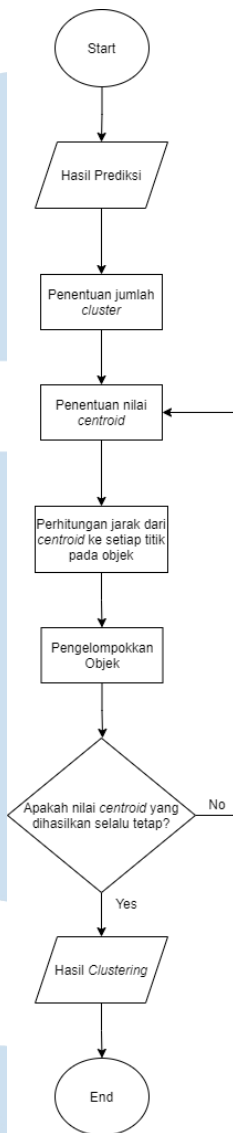
Di sisi lain dalam hal *clustering* akan dilakukan dengan menggunakan metode *K-Means*. Pemilihan metode *K-Means* ini karena dalam metode ini penentuan atributnya dapat ditentukan secara pasti dan efisien dalam menangani *dataset* yang besar. Untuk lebih jelasnya dapat dijelaskan pada Tabel 3.7 di bawah ini.

Tabel 3.7 Perbandingan K-Means dan DB Scan

Faktor	K-Means	DB Scan
Pembeda		
Penentuan Atribut	Penentuan atribut k dapat ditentukan dengan menggunakan <i>elbow method</i> [48]	Penentuan atribut Epsilon sulit ditentukan untuk memperoleh hasil <i>clustering</i> yang terbaik [49]
Penanganan Dataset	Efisien dalam menangani dataset yang besar	Kurang efisien dalam menangani dataset yang besar [50]

Secara garis besar Gambar 3.4 berikut merupakan langkah-langkah penggunaan metode *K-Means* dalam sub-metode *Data Mining*.





Gambar 3.4 Langkah Metode K-Means

Hasil prediksi yang sudah didapatkan sebelumnya pada metode *Holt-Winters* selanjutnya dijadikan *input* pada metode *K-Means*. Setelah itu akan dilakukan penentuan jumlah *cluster* atau atribut *k* menggunakan *elbow method*. Setelah itu dilanjutkan dengan penentuan nilai *centroid* atau titik pusat *cluster*. Langkah selanjutnya adalah perhitungan jarak dari titik pusat *cluster* ke masing-masing titik pada objek, perhitungan ini dilakukan dengan menggunakan *Euclidean Distance*. Setelah itu dilakukan pengelompokan objek ke dalam bentuk *cluster-cluster*. Langkah selanjutnya adalah pengecekan apakah nilai *centroid* yang dihasilkan selalu

sama atau tidak, apabila tidak maka langkah akan diulang lagi mulai dari penentuan nilai *centroid* namun apabila nilai *centroid* yang dihasilkan selalu sama maka hasil *cluster* sudah ditemukan. Hasil *cluster* yang ditemukan merupakan *output* dari penggunaan metode *K-Means* dalam penelitian ini.

3.2.5 Evaluation/ Interpretation

Evaluasi dilakukan untuk mengukur tingkat akurasi dari pemodelan prediksi yang sudah dibuat dengan menggunakan metode *Holt-Winters*. Proses evaluasi dilakukan dengan menggunakan metode MAPE. Pemilihan penggunaan metode MAPE ini karena metode MAPE akan memberikan informasi mengenai kesalahan hasil peramalan dalam periode tertentu apakah terlalu tinggi atau terlalu rendah dalam bentuk persentase. Sesuai dengan penelitian ini, metode MAPE akan memperhatikan apakah hasil peramalan lebih kecil atau lebih besar daripada kenyataan dan MAPE akan memberikan *output* dalam bentuk persentase sehingga lebih mudah dimengerti. Untuk lebih jelasnya dapat dilihat pada Tabel 3.8 di bawah ini.



Tabel 3.8 Perbandingan Metode MAPE dan MAD

Faktor Pembeda	<i>Mean Absolute Percentage Error (MAPE)</i>	<i>Mean Absolute Deviation (MAD)</i>
Kegunaan	Mengukur taraf kesalahan <i>output</i> prediksi saat variabel peramalan merupakan faktor yang krusial pada mengevaluasi taraf akurasi menurut prediksi yang dihasilkan .	Mengukur tingkat kesalahan hasil prediksi dalam unit ukuran yang sama seperti data aslinya.
Tingkat Akurasi Hasil	Tingkat akurasi hasil prediksi dalam periode tertentu dinyatakan dalam bentuk persentase, di mana persentase tersebut memberikan informasi mengenai kesalahan dari hasil peramalan apakah terlalu tinggi atau tidak.	Tingkat akurasi hasil peramalan tergolong sangat kecil karena tidak memperhatikan apakah hasil peramalan lebih kecil atau lebih besar daripada kenyataan.

Sumber: [51]

Nilai persentase yang dihasilkan oleh metode MAPE akan dijadikan acuan untuk menentukan apakah pemodelan prediksi mengenai total kasus aktif Covid-19 dengan menggunakan metode *Holt-Winters* layak atau tidak. Terdapat skala interval yang membagi MAPE menjadi beberapa kategori berdasarkan persentase yang dihasilkan. Tabel 3.9 berikut akan menjelaskan tentang skala nilai MAPE.

Tabel 3.9 Tabel Skala Nilai MAPE

Range MAPE	Arti Nilai
< 10%	Kemampuan model peramalan sangat baik
10 - 20%	Kemampuan model peramalan baik
20 - 50 %	Kemampuan model peramalan layak
>50%	Kemampuan model peramalan buruk

Sumber: [52]

Dapat dilihat pada Tabel 3.9 bahwa apabila persentase nilai MAPE semakin kecil maka hasil pemodelan prediksi yang dihasilkan semakin baik dan sebaliknya semakin besar persentase nilai MAPE maka hasil pemodelan prediksi yang dihasilkan semakin buruk. Berdasarkan rentang skala pada Tabel 3.9 dapat diketahui bahwa metode prediksi yang digunakan tidak layak untuk digunakan ketika nilai MAPE lebih besar dari 50%.

Interpretasi akan dilakukan berdasarkan hasil *clustering* dengan menggunakan metode *K-Means*. Bentuk interpretasi yang dilakukan adalah dengan membuat visualisasi *dashboard* yang menentukan apakah pengunjung dari Pusat Kebugaran X boleh datang langsung ke lokasi atau tidak berdasarkan jumlah kasus aktif Covid-19 dalam periode waktu tertentu. Pembuatan visualisasi *dashboard* akan dilakukan dengan menggunakan bantuan *tools Tableau*. Pemilihan *tools Tableau* ini karena dapat menangani *dataset* yang besar. Selain itu *tools Tableau* ini juga dapat mengakses sumber data dari berbagai *database* dan *server* yang jauh lebih banyak daripada *Power BI*. Sesuai dengan penelitian ini yang bertujuan untuk menghasilkan laporan dan visualisasi data yang dinamis dan bukan untuk menganalisis data bisnis. Untuk penjelasan lebih lengkap dapat dilihat pada Tabel 3.10.

Tabel 3.10 Tabel Perbandingan Tableau dan Power BI

Faktor Pembeda	Tableau	Power BI
Pengertian	<i>Tools</i> untuk menganalisis data dengan tujuan untuk menghasilkan laporan maupun visualisasi data dengan fleksibilitas yang tinggi	<i>Tools</i> untuk menganalisis data bisnis dengan tujuan untuk memperoleh informasi bisnis dari data tersebut
Sumber Data	Memiliki akses ke berbagai sumber <i>database</i> maupun server yang jauh lebih banyak daripada <i>Power BI</i>	Akses ke <i>database</i> dan <i>server</i> terbatas dan jauh lebih sedikit jika dibandingkan dengan <i>Tableau</i>
<i>Performance</i>	Dapat menangani <i>dataset</i> yang sangat besar dengan kinerja yang sangat baik	Volume data yang dapat ditangani terbatas yaitu 3.500 data

Sumber: [53]

3.3 Variabel Penelitian

Dalam penelitian ini terdapat variabel independen dan juga variabel dependen. Di mana variabel independen akan mempengaruhi nilai dari variabel dependen.

3.3.1 Variabel Dependen

Variabel dependen atau variabel terikat merupakan variabel yang nilainya dipengaruhi oleh variabel independen atau variabel bebas. Dalam penelitian ini variabel dependen saat metode prediksi dengan *Holt-Winters* adalah total kasus aktif Covid-19/ *Total Active Cases*. Pada saat *clustering* dengan metode *K-Means* yang menjadi variabel dependen adalah variabel *cluster*.

3.3.2 Variabel Independen

Variabel independen atau variabel bebas merupakan variabel yang mempengaruhi nilai daripada variabel dependen. Dalam penelitian ini

variabel independen saat metode prediksi dengan *Holt-Winters* adalah variabel *Date*. Pada saat *clustering* dengan metode *K-Means* yang menjadi variabel independen adalah variabel *Date* dan hasil prediksi total kasus aktif Covid-19 yang sudah dilakukan dengan menggunakan metode *Holt-Winters*.

3.4 Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan adalah teknik *dataset* statistik yaitu metode pengumpulan data statistik dengan menggunakan data yang sudah ada [54]. Data yang akan digunakan dalam penelitian ini adalah data mengenai jumlah kasus harian Covid-19 di beberapa provinsi di Indonesia antara lain provinsi DKI Jakarta, Banten dan Jawa Barat. Untuk pengumpulan data Covid-19 dari provinsi DKI Jakarta pada penelitian ini didapatkan dan diunduh dari *website* <https://corona.jakarta.go.id/id>, sedangkan untuk pengumpulan data Covid-19 dari provinsi Banten didapatkan dan diunduh dari *website* <https://dinkes.bantenprov.go.id/> dan untuk data Covid-19 dari provinsi Jawa Barat pada penelitian ini didapatkan dan diunduh dari *website* <https://pikobar.jabarprov.go.id/data>. Data yang didapatkan dalam bentuk file *excel*.

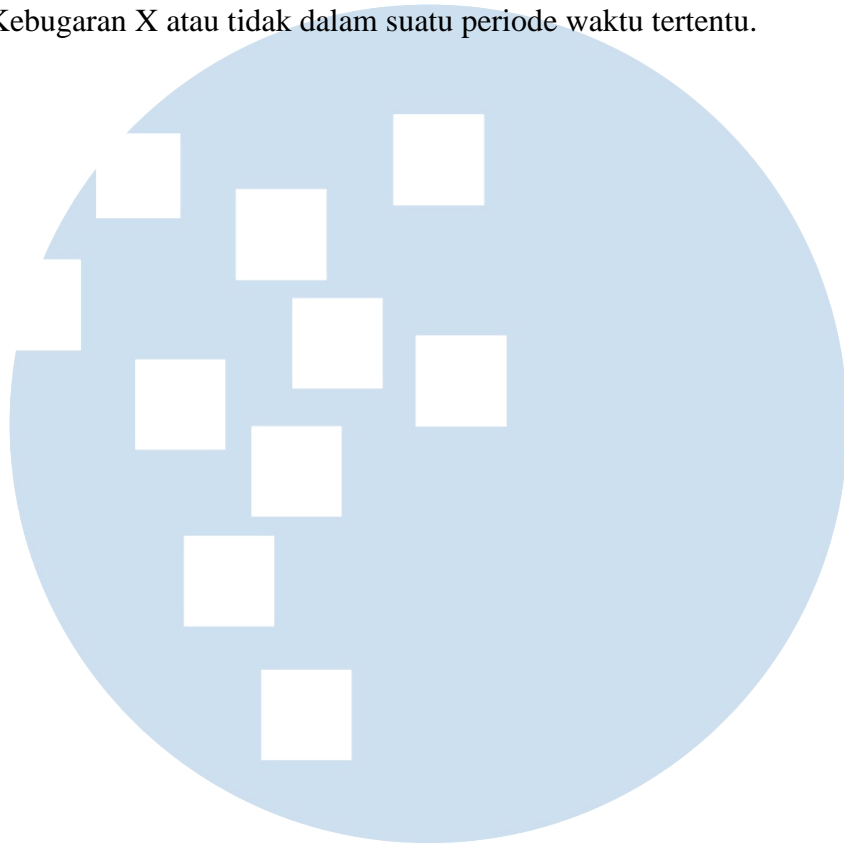
3.5 Teknik Pengambilan Sampel

Teknik pengambilan sampel yang digunakan adalah teknik *purposive sampling*. Teknik ini merupakan pemilihan sampel berdasarkan populasi dengan karakteristik tertentu yang mungkin berkaitan erat dengan populasi yang telah diketahui sebelumnya. [55]. Di mana dalam penelitian ini pengambilan sampel data Covid-19 hanya melibatkan populasi yang memiliki karakteristik khusus atau mengalami gejala terjangkit Covid-19 serta orang-orang yang sudah terjangkit Covid-19.

3.6 Teknik Analisis Data

Teknik analisis data yang digunakan adalah teknik analisis statistik inferensial. Teknik ini merupakan teknik analisis data yang mengharuskan peneliti untuk membuat kesimpulan dari proses analisis yang sudah dilakukan [56]. Di mana penelitian ini memiliki kesimpulan dari hasil analisis data berupa visualisasi

dashboard yang memberikan keputusan apakah seseorang boleh mengunjungi Pusat Kebugaran X atau tidak dalam suatu periode waktu tertentu.



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA