

BAB II

KAJIAN LITERATUR

2.1 Polusi Udara

Polusi udara merupakan salah satu permasalahan lingkungan yang memiliki pengaruh penting terhadap kehidupan manusia dan ekosistem [9]. Secara umum, polusi udara berbentuk partikel (debu, aerosol, timah hitam) dan gas (CO, NO_x, SO_x, H₂S, hidrokarbon). Udara yang tercemar dengan partikel dan gas ini dapat menyebabkan gangguan kesehatan yang berbeda tingkatannya dan jenisnya, tergantung dari macam, ukuran dan komposisinya. Gangguan tersebut terutama terjadi pada fungsi vital dari organ tubuh seperti paru-paru dan pembuluh darah atau menyebabkan iritasi pada mata dan kulit [10].

Penelitian ini dilakukan untuk mengetahui indeks pencemaran polusi udara di kota Jakarta, Indeks Standar Pencemar Udara (ISPU) berdasarkan Peraturan Pemerintah Republik Indonesia No.41 tahun 1999 yaitu suatu angka yang tidak mempunyai satuan yang dapat menggambarkan kondisi mutu udara ambien di lokasi tertentu, yang umumnya akan berdampak pada kesehatan manusia, nilai estetika dan makhluk hidup [3]. Polusi udara berasal dari berbagai sumber seperti pembakaran batu bara, pembakaran BBM pada sarana transportasi (darat, laut dan udara), pembakaran pada proses industri dan pengolahan limbah domestik, serta zat kimia yang tanpa disadari langsung diemisikan ke udara oleh kegiatan manusia [3]. Udara di alam tidak pernah ditemukan bersih tanpa polutan sama sekali, beberapa gas seperti Sulfur Dioksida (SO₂), Nitrogen Dioksida (NO₂), Karbon Monoksida (CO) dan Ozon (O₃) selalu dibebaskan ke udara sebagai efek dari proses alami seperti pembusukan sampah tanaman, aktivitas vulkanik, kebakaran hutan dan sebagainya.

2.2 Forecasting (Peramalan)

Forecasting (Peramalan) adalah sebuah proses untuk memperkirakan suatu kejadian yang akan terjadi dimasa depan dan dilakukan dengan adanya data historis yang dikalkulasikan dengan metode tertentu untuk memperkirakan sebuah prediksi yang akan terjadi dimasa depan [8]. *Forecasting* dilakukan bertujuan untuk menentukan ketepatan dan kekuatan yang diperlukan serta mengkaji kebijakan yang berlaku saat ini dan dimasa lalu serta melihat sejauh mana pengaruh di masa depan. Kegunaan dari dilakukannya *forecasting* (peramalan) dapat dilihat pada saat pengambilan keputusan, Keputusan yang baik adalah keputusan yang didasarkan oleh

pertimbangan apa yang akan terjadi Ketika keputusan tersebut dijalankan [15]. Teknik *forecasting* terbagi atas 2 (dua) kategori utama berdasarkan sifatnya yaitu:

1. Metode *forecasting* (peramalan) kualitatif atau teknologis

Peramalan kualitatif merupakan peramalan yang didasarkan atas data kualitatif di masa lalu. Peramalan yang dilakukan tersebut ditentukan berdasarkan pemikiran yang bersifat intuisi, pendapat dan pengetahuan dari orang menyusunnya.

2. Metode *forecasting* (peramalan) kuantitatif

Peramalan kuantitatif merupakan peramalan yang didasarkan atas data kuantitatif pada masa lalu. Peramalan dengan metode kuantitatif diperkirakan lebih efektif dalam pengaplikasiannya dibandingkan dengan metode kualitatif karena ketepatan akurasi dari hasil pengukuran yang didapat berdasarkan data-data yang telah dikumpulkan sehingga peramalan yang dilakukan secara kuantitatif memiliki peluang kemungkinan terjadi [16].

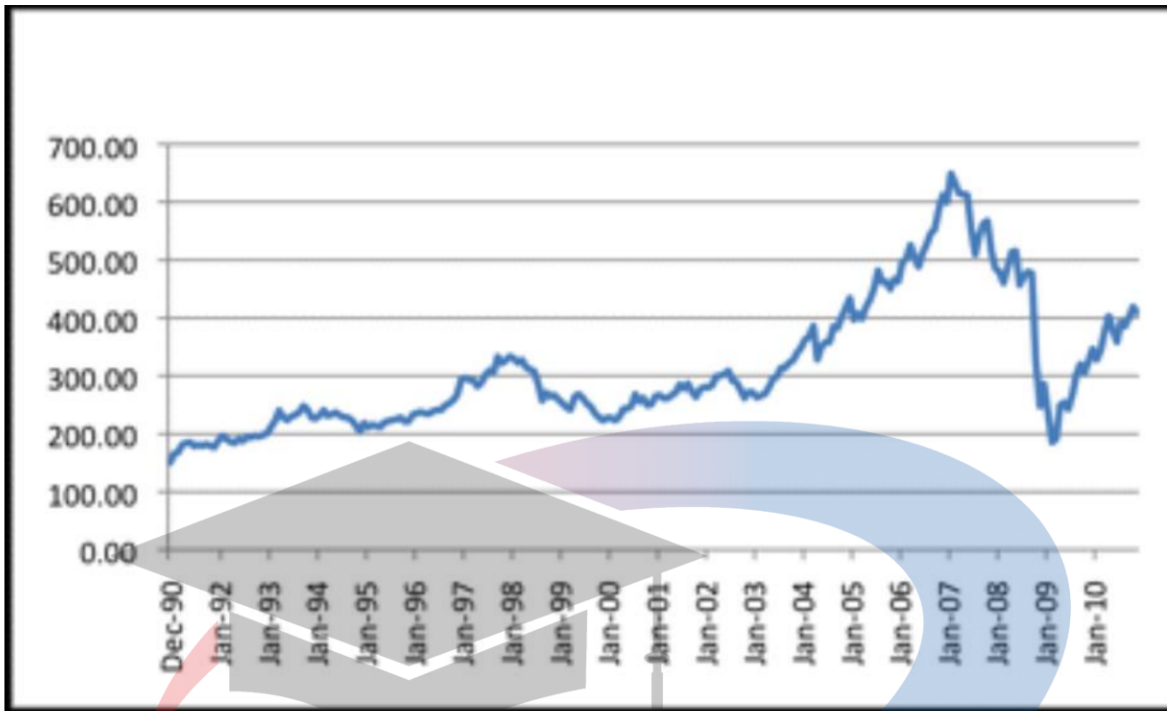
2.3 Data Deret Waktu

Deret waktu (*time series*) metode yang dipergunakan untuk menganalisis serangkaian data yang merupakan fungsi waktu [17]. Data yang digunakan adalah data-data yang sudah dikumpulkan secara periodik berdasarkan urutan waktu. Untuk menentukan metode peramalan pada metode deret waktu perlu diketahui dari pola data tersebut sehingga peramalan data dapat dilakukan dengan metode yang sesuai. Data *time series* dapat dicatat berdasarkan periode waktu harian, mingguan, bulanan, tahunan, ataupun periode waktu tertentu dalam rentang waktu yang sama [18]. Dalam Deret waktu (*time series*) suatu periode waktu dipengaruhi oleh nilai pada periode sebelumnya. Dan kemudian Analisis data *time series* dapat dilakukan peramalan (*forecasting*) dimasa yang akan datang.

Peramalan pada data *time series* berguna untuk kepentingan banyak pihak, selain dapat memprediksi nilai pada suatu data, peramalan pada data *time series* juga dapat menentukan suatu kebijakan tertentu tergantung dari ukuran besar atau kecilnya nilai dari prediksi tersebut diperoleh.

2.3.1. Plot Runtun Waktu

Diagram deret waktu (*time series*) atau yang disebut juga dengan grafik deret waktu atau plot deret waktu adalah alat visualisasi data yang mengilustrasikan titik data pada interval waktu yang berurutan [19].



Gambar 2.1 Contoh plot time series

Berikut adalah jenis-jenis dari plot yang digunakan untuk analisis deret waktu :

1. Plot Waktu

Plot waktu adalah plot yang berbentuk garis yang menunjukkan evolusi *series* dari waktu ke waktu. Plot ini dapat digunakan sebagai awalan dalam menganalisis untuk mendapatkan pemahaman dasar pada data yang sedang dianalisis [20].

2. Plot Musiman

Plot Musiman adalah plot yang hampir sama dengan plot waktu bedanya adalah data yang digunakan pada plot musiman adalah data dari masing-masing musim yang ditentukan [20].

3. Plot Musiman Polar

Plot ini merupakan variasi dari plot musiman, dengan perbedaannya plot musiman polar menggunakan koordinat kutub.

4. Plot Dekomposisi Deret Waktu

Plot Dekomposisi Deret Waktu adalah plot yang mengkombinasikan komponen level, tren, musiman, dan gangguan.

5. Plot Autokorelasi

Dalam plot autokorelasi ada 2 jenis fungsi yang digunakan yaitu:

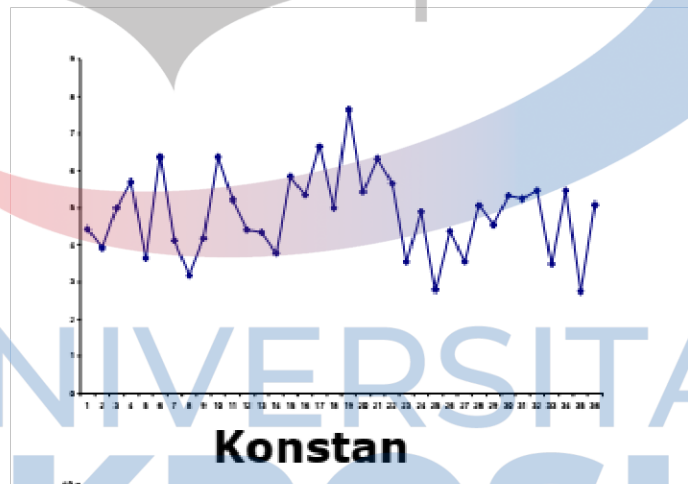
- a. Fungsi Autokorelasi (ACF) merupakan fungsi yang menunjukkan nilai koefisien korelasi antara deret. ACF mempertimbangkan semua komponen deret waktu sambil menemukan korelasi dan nilai lag-nya.
- b. Fungsi Autokorelasi Parsial (PACF) merupakan fungsi kebalikan dari ACF. Fungsi Autokorelasi Parsial melihat korelasi antara residual dan nilai berikutnya. Kemudian menghapus variasi yang sudah ditemukan sebelum menemukan korelasi berikutnya.

2.3.2. Pola Runtun Waktu

Terdapat beberapa pola dalam runtun waktu, yaitu [21]:

1. Konstan (*Horinzontal*)

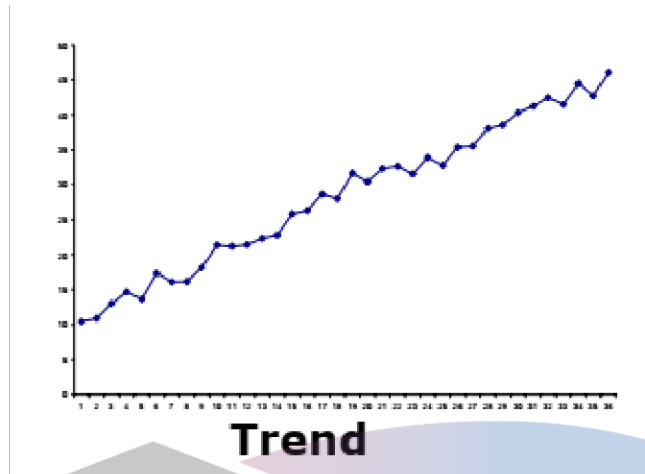
Pola data yang akan terjadi apabila data berfluktualisasi disekitar nilai rata-rata konstan (stasioner terhadap nilai rata-ratanya).



Gambar 2.2 Pola Data Horizontal.

2. Tren (*Tren*)

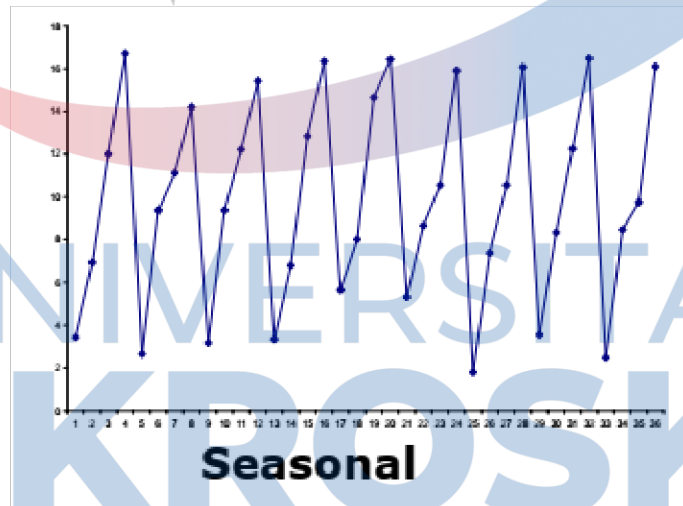
Pola tren terjadi apabila terjadi kenaikan atau penurunan dalam kurun waktu dalam data. Pola ini dipengaruhi oleh faktor budaya, perubahan populasi dan perubahan pendapatan.



Gambar 2.3 Pola Data Tren

3. Musiman

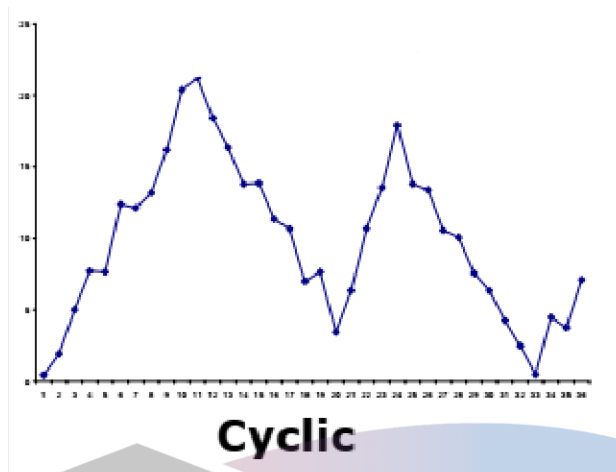
Pola data yang akan terjadi apabila ada rangkaian data yang dipengaruhi oleh faktor musiman. Pola ini muncul secara reguler setiap tahun yang biasa disebabkan oleh iklim, kebiasaan (mempunyai pola tetap dari waktu ke waktu). Contohnya adalah kegiatan pada hari-hari besar dalam kalender dan hari libur keluarga



Gambar 2.4 Pola Data Musiman

4. Siklus

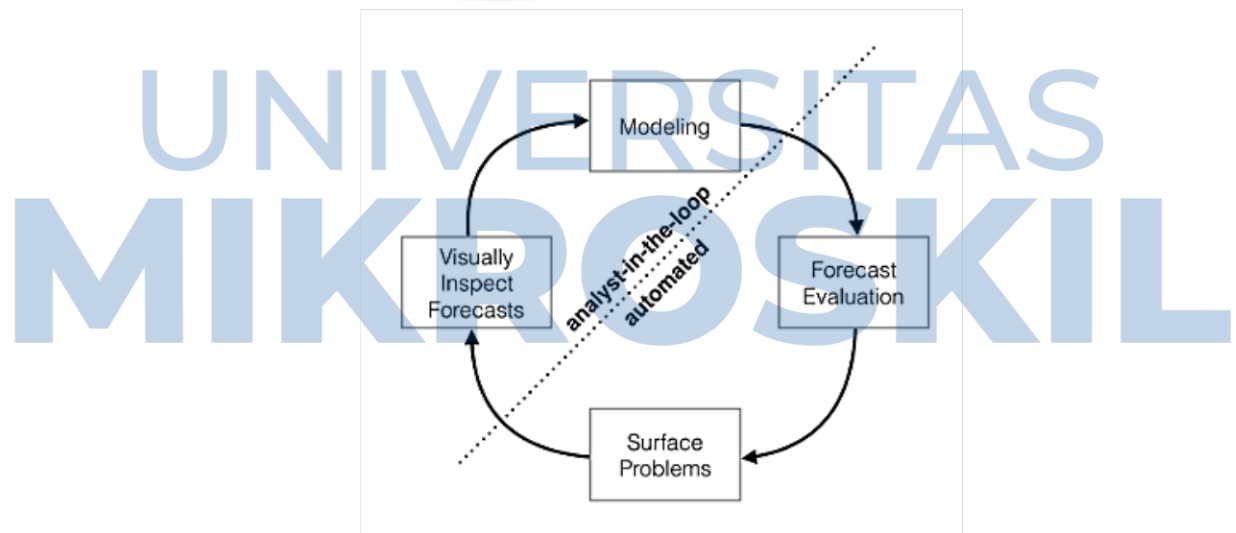
Pola data yang terjadi apabila ada data yang dipengaruhi oleh fluktuasi kegiatan ekonomi yang mempunyai jarak muncul dan sebab yang berbeda-beda.



Gambar 2.5 Pola Data Siklus

2.4 Prophet facebook

Metode *prophet facebook* merupakan model yang secara otomatis dapat mendeteksi perubahan poin dalam runtun waktu yang dapat disesuaikan dengan cara intuitif untuk analisis yang belum berpengalaman. Model ini menghasilkan perkiraan yang akurat dan reliable yang memiliki karakteristik tren yang merupakan kurva pertumbuhan non-linier, di mana tren mencapai batas alami atau jenuh, perubahan tren historis, pengamatan setiap jam, harian, atau mingguan dengan setidaknya beberapa bulan (satu tahun) dari historis, dan beberapa musiman “skala manusia” yang kuat [22].



Gambar 2.6 Skematik Pendekatan forecasting at scale [8]

Metode penelitian yang dilakukan dalam memprediksi tingkat polusi udara sebagai studi kasus pada penelitian ini, digunakan metode algoritma prophet yang mampu menghasilkan prediksi

dengan menggunakan metode *Time Series Forecasting* yang dapat diurai (*decomposable time series*) dengan tiga komponen mode utama yaitu tren (*trends*), musiman (*seasonality*), dan hari libur (*Holidays*) yang dikombinasikan menjadi persamaan berikut [23]:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (1)$$

Dimana komponen dari persamaan (1), terdiri dari :

- $g(t)$: fungsi *trend* yang memodelkan perubahan non-periodik pada nilai data time series
- $s(t)$: mewakili perubahan periodic (musiman)
- $h(t)$: mewakili dampak dari hari libur
- ϵ_t : mewakili perubahan istimewa yang tidak diakomodasi oleh model.

Metode *Prophet* didukung oleh dua model tren yaitu pertumbuhan jenuh (*saturating growth*) dan linear sepotong-sepotong (*piecewise linear*). Pertumbuhan jenuh (*saturating growth*) dinyatakan melalui persamaan dibawah ini [23]:

$$g(t) = \frac{c}{1 + \exp(-k(t - m))} \quad (2)$$

Dimana komponen dari persamaan (2), terdiri dari :

- c : merupakan kapasitas pembawa (*carrying capacity*)
- k : tingkat pertumbuhan (*growth rate*)
- m : parameter penyeimbang (*offset*).

Kemudian linier sepotong-sepotong dinyatakan melalui persamaan dibawah ini :

$$g(t) = \frac{c}{1 + \exp\left(-\left(k + a(t) \uparrow \delta\right)\left(t - \left(m + a(t) \uparrow \gamma\right)\right)\right)} \quad (3)$$

Dimana komponen dari persamaan (3), terdiri dari :

- k : tingkat pertumbuhan
- δ : mewakili penyesuaian (*adjustments*)
- m : parameter penyeimbang.

Metode *Prophet* didukung oleh model musiman dengan menggunakan deret *fourier* untuk fleksibilitas yang dinyatakan melalui persamaan dibawah ini [23]:

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{p}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right) \quad (4)$$

Dimana komponen dari persamaan (4), terdiri dari :

P : periode waktu

Metode *Prophet* didukung oleh model hari libur yang dinyatakan melalui persamaan dibawah ini:

$$Z(t) = [1(t \in D_1), \dots, 1(t \in D_1)] \quad (5)$$

Dimana komponen dari persamaan (5), terdiri dari :

D : set dari tanggal liburan di masa lalu dan di masa depan

t : waktu saat sedang liburan

Dalam mencapai hasil prediksi yang diinginkan dengan model *Prophet facebook* dilakukan dengan bantuan *software R programming*. Di satu sisi, tren mengutarakan bagaimana keseluruhan runtun waktu (*time series*) tumbuh dan bagaimana itu diharapkan tumbuh dimasa depan [22].

2.5 Support vector Regression (SVR)

SVR merupakan penerapan SVM (*Support Vector Machine*) yang digunakan untuk kasus regresi yang outputnya berupa bilangan riil atau kontinyu. Dalam kata lain SVM untuk kasus regresi disebut SVR. Tujuan algoritma SVR adalah untuk menemukan sebuah fungsi $f(x)$ sebagai suatu *Hyperplane* (garis pemisah) terbaik yang sesuai dengan semua input data dengan sebuah error ϵ dan membuat ϵ setipis mungkin [24]. Metode SVM digunakan untuk kasus klasifikasi dan dapat dikembangkan ke dalam regresi dan metode peramalan (*Time Series*). Misalnya terdapat 1 data training , (x_i, y_i) , $i = 1, \dots, 1$ dengan data input $x = \{x_{(1)} \dots, x_{(1)}\} \subseteq \mathbb{R}^N$ dan $y = \{y_{(1)} \dots, y_{(1)}\} \subseteq \mathbb{R}$ dan 1 adalah banyaknya data training [25]. Fungsi regresinya adalah sebagai berikut:

$$F(x) = w \phi(x) + b \quad (6)$$

Dimana komponen dari persamaan diatas terdiri dari:

w : vektor pembobot

$\varphi(x)$: fungsi yang memetakan x dalam suatu dimensi
 b : bias

Dalam meminimalkan fungsi resiko (*risk function*) dari persamaan diatas dapat diestimasi melalui cara sebagai berikut:

$$\min_{w, \xi_i, \xi_i^*} \frac{1}{2} ||w||^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (7)$$

Dengan kendala

$$\begin{aligned} y_i &\leq f(\mathbf{x}_i) + \varepsilon + \xi_i^* \\ y_i &\geq f(\mathbf{x}_i) - \varepsilon - \xi_i \\ \xi_i^*, \xi_i &\geq 0 \end{aligned} \quad (8)$$

Konstanta $C > 0$, berfungsi untuk menentukan *trade off* (tawar menawar) antara ketipisan fungsi f dan batas atas deviasi yang lebih kecil dari ε , yang masih bisa di toleransi. C adalah parameter untuk menentukan berapa besar deviasi dari parameter epsilon yang masih bisa ditoleransi, yaitu semakin besar nilai C maka semakin besar nilai toleransi yang masih diijinkan. Parameter epsilon (ε) berfungsi untuk mengontrol lebar zona regresi dalam pemrosesan data. Semakin besar nilai epsilon maka estimasi regresi semakin datar mendekati regresi linier.

Koefisien epsilon (ε) dalam algoritma SVR adalah ekuivalen dengan tingkat akurasi data training. Nilai ε yang kecil berhubungan dengan nilai tinggi pada variabel *slack* dengan akurasi tinggi. Sebaliknya jika nilai ε yang besar maka berhubungan dengan nilai dari variabel *slack* dengan akurasi yang rendah. Penambahan variabel *slack* ini berguna untuk mengatasi masalah pembatas margin yang tidak layak dalam problem optimasi.

2.5 Evaluasi Model

Untuk mengukur kinerja algoritma forecasting dengan mengevaluasi perbandingan antara hasil prediksi dan hasil aktualnya dapat dilakukan dengan cara berikut :

1. Mean Absolute Percentage Error (MAPE)

Mean absolute percentage error (MAPE) adalah nilai rata-rata perbedaan absolut yang ada diantara nilai dari prediksi dan nilai aktual sebagai hasil persenan dari nilai aktual [26]. Nilai MAPE dapat dihitung dengan persamaan berikut

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i^{\wedge} - y_i}{y_i} \right| \quad (9)$$

Dimana :

y_i^{\wedge} : Nilai prediksi

y_i : Nilai aktual

n : Jumlah data

2. Root Mean Square Error (RMSE)

Mean Square Error (RMSE) adalah perhitungan dengan cara menjumlahkan semua kuadrat kesalahan prediksi. Kemudian membagi jumlah tersebut dengan banyaknya waktu prediksi [27]. Nilai RMSE dapat dihitung dengan persamaan berikut :

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{y}_i - y_i)^2} \quad (10)$$

Dimana :

\tilde{y}_i : Nilai hasil prediksi

y_i : Nilai aktual

n : Jumlah data

3. Mean Absolute Error (MAE)

Mean absolute error (MAE) adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan model peramalan [28]. Nilai MAE dapat dihitung dengan persamaan berikut :

$$MAE = \frac{1}{n} \sum_{i=1}^n |\tilde{y}_i - y_i| \quad (11)$$

dimana :

\tilde{y}_i : Nilai hasil prediksi

y_i : Nilai aktual

n : Jumlah data

4. Mean Square Error (MSE)

Mean Square Error (MSE) adalah hasil pengurangan antara nilai aktual dan prediksi yang telah dikuadratkan dan kemudian dilakukan penjumlahan terhadap hasil tersebut [29]. Nilai MSE dapat dihitung dengan persamaan berikut :

$$MSE = \frac{1}{n} \sum_{i=1}^n (\tilde{y}_i - y_i)^2 \quad (12)$$

dimana :

\tilde{y}_i : Nilai hasil prediksi

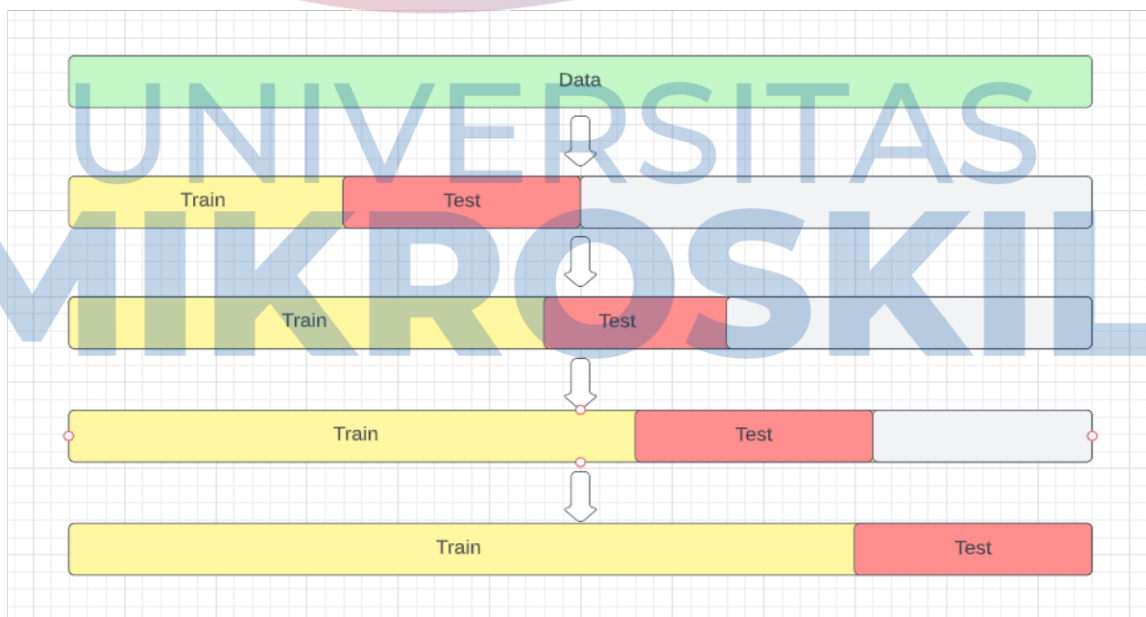
y_i : Nilai aktual

n : Jumlah data

2.6 Validasi Silang (*Cross Validation*)

Proses evaluasi kinerja model dilakukan dengan membagi data menjadi dua bagian yakni data *train* dan data *test*. Namun hal demikian dapat menyebabkan terjadinya *overfitting* [30].

Cara alternatif untuk evaluasi kinerja model dan mencegah terjadinya *overfitting* adalah dengan validasi silang (*cross validation*). Metode yang digunakan untuk *cross validation* deret waktu adalah validasi silang bergulir. Dengan memulai sebagian kecil data *test* kemudian periksa akurasi pada sebagian data setelahnya yang digunakan sebagai peramalan. Data digunakan dalam peralamalan kemudian akan digabungkan dengan data *train* sebelumnya dan digunakan kembali sebagai data *train* selanjutnya kemudian periksa kembali akurasi pada sebagian data setelahnya [30]. Lihat contoh pembagian alur kerja validasi silang pada gambar 2.7.



Gambar 2.7 Contoh pembagian data Cross Validation

2.7 Literatur Review

Dalam penelitian ini, terdapat beberapa studi literatur terkait dari penelitian sebelumnya sebagai acuan sebagai referensi dalam penelitian ini. Dalam tabel 2.1 terdapat hasil tinjauan pustaka dari penelitian terdahulu.

Tabel 2.1 Studi Literatur

No	Referensi	Models	Kontribusi	Area aplikasi	Hasil
1	[6]	<i>Prophet</i> <i>Facebook</i>	Prediksi Jumlah Mahasiswa Baru selama 5 tahun (2020 – 2024)	Pendidikan	Hasil penelitan ini diperoleh nilai MAPE sebesar 0,04327568%.
2	[7]	<i>Prophet</i> <i>Facebook</i>	Prediksi Harga pangan di kota Bandung selama 1 kuartal	Ekonomi	Hasil penelitian ini diperoleh persentasi rata-rata akurasi sebesar 84,243% untuk model non-linear dan 81,379% untuk model linear.
3	[23]	<i>Prophet</i> <i>Facebook</i>	Prediksi harga saham bank BCA di periode 2017 sampai 2020	Ekonomi atau Keuangan	Hasil penelitian ini diperoleh nilai MAPE sebesar 5.47%.
4	[31]	<i>Prophet</i> <i>Facebook</i>	Prediksi puncak wabah Covid-19 di Indonesia di bulan maret sampai oktober 2020	Kesehatan	Hasil penelitian ini adalah prediksi Covid-19 di Indonesia akan terus naik dan prediksi pada tanggal 22 oktober 2020 terdapat 375.395 orang yang terinfeksi Covid-19 di Indonesia.

No	Referensi	Models	Kontribusi	Area aplikasi	Hasil
5	[8]	<i>Prophet</i> <i>Facebook</i> dan ARIMA	Analisis komparatif antara <i>Prophet</i> dan ARIMA dalam Pendaftaran mahasiswa baru pada periode 2017-2019	Pendidikan	Hasil penelitian ini menunjukkan prediksi yang lebih baik menggunakan metode <i>Prophet</i> dengan tingkat akurasi RSME berturut-turut 0.135, 0.688, 0.166, 0.291, dan 0.232.
6	[32]	SVR dan ARIMA	Prediksi jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin.	Transportasi	Hasil ini menunjukkan prediksi model SVR yang lebih baik dengan MAPE dengan data training dan data testing berdasarkan MAPE sebesar 7,5% dan 11,3%.
7	[33]	Deret Fourier dan SVR	Prediksi Harga Ekspor Non migas di Indonesia	Ekonomi	Dari hasil prediksi ini menunjukkan bahwa model SVR memprediksi dengan lebih akurat dengan MAPE masing-masing sebesar 9,29% dan 16,03%.