

## BAB II LANDASAN TEORI

### 1.1. Tinjauan Pustaka

Beberapa penelitian terdahulu yang dijadikan referensi dalam penelitian ini dapat dilihat pada Tabel 2.1:

Tabel 2.1 Tinjauan Pustaka

1	Judul Penelitian	Stock Market Price Forecasting using ARIMA vs ANN; A Case study from CSE
	Tahun Penelitian	2022
	Metode Penelitian	ARIMA dan ANN
	Penulis	G.W.R.I Wijesinghe
	Permasalahan	Berfokus pada bursa efek colombo (CSE) memiliki 290 perusahaan yang mewakili 20 group industri, investor kesulitan dalam menganalisa dan membuat sebuah keputusan
	Kesimpulan	Berdasarkan penelitian diketahui dengan menunjukkan perbandingan antara ARIMA dan ANN bahwa metode ANN lebih unggul dari ARIMA .nilai MAPE metode ANN 0.17833 dan metode ARIMA 0.4672206
2	Judul Penelitian	Stock Price Prediction Using Time Series, Econometric, Machine Learning, and Deep Learning Models
	Tahun Penelitian	2021
	Metode Penelitian	ARIMA Random Forest, RNN dan LSTM
	Penulis	Ananda Chatterje
	Permasalahan	Peneliti ingin membangun model yang kuat untuk memprediksi harga saham di 3 sektor berbeda TI,Perbankan dan Kesehatan

	Kesimpulan	Berdasarkan penelitian dengan membandingkan metode machine learning dan deep learning diketahui bahwa metode ARIMA lebih baik dari metode machine learning lainnya sesuai perbandingan dan metode LSTM lebih baik dari kategori metode deep learning lainnya.
3	Judul Penelitian	Perbandingan Metode Holt-Winter Exponential Smoothing dan Weighted Fuzzy Integrated Time Series Untuk Memprediksi LQ45
	Tahun Penelitian	2021
	Metode Penelitian	Holt-Winter Exponential Smoothing
	Penulis	Diah Nur Hastuti
	Permasalahan	Penelitian ini mengenai prediksi harga saham, berdasarkan kecenderungan trend pergerakan harga sama dimasa lalu
	Kesimpulan	Berdasarkan hasil penelitian diketahui metode Weighted Fuzzy Integrated Time Series merupakan metode terbaik untuk memprediksi data LQ45 periode Januari, Februari dan Maret 2021 karena menghasilkan nilai MAPE lebih kecil
4	Judul Penelitian	Perbandingan Metode ARIMA dan Double Exponential Smoothing pada Peramalan Harga Saham LQ45 Tiga Perusahaan dengan Nilai <i>Earning Per Share</i> (EPS) Tertinggi
	Tahun Penelitian	2017
	Metode Penelitian	ARIMA, Double Exponential Smoothing
	Penulis	Irma Fitria
	Permasalahan	Perubahan harga saham dari waktu ke waktu menyulitkan investor dalam melakukan analisis secara fundamental. Investor perlu mengetahui informasi mengenai harga saham kedepan . Peneliti

		menggunakan Metode ARIMA dan Exponential Smoothing untuk melakukan peramalan
	Kesimpulan	Berdasarkan hasil penelitian menggunakan data saham periode 2011 sampai 2015 diketahui Metode Exponential Smoothing memiliki nilai MAPE lebih kecil dibandingkan ARIMA
5	Judul Penelitian	Stock Price Prediction Using Facebook Prophet and ARIMA models
	Tahun Penelitian	2021
	Metode Penelitian	FB-Prophet dan ARIMA
	Penulis	Anusha Garlapati
	Permasalahan	Dengan pandemi yang sedang berlangsung COVID-19 patronase seperti pasar saham, tekstil menjadi surut, diperlukan ramalah saham untuk mengetahui harga saham dimasa mendatang
	Kesimpulan	Berdasarkan hasil penelitian dengan menggunakan data saham periode 2012-2020 menunjukkan hasil bahwa metode ARIMA (2,1,2) lebih baik dalam memprediksi dibandingkan dengan metode FB-Prophet

Tabel 2. 1

## 1.2. Perbedaan dengan penelitian yang dilakukan

Berdasarkan tinjauan pustaka yang dilakukan oleh penulis, maka perbedaan antara penelitian terdahulu dan penelitian yang sedang dilakukan adalah:

1. Data yang digunakan adalah data harga saham BBRI dengan periode 2012 – 2022
2. Penulis menggunakan metode ARIMA, Exponential Smoothing, FB-Prophet, dan Transformer.
3. Penulis menggunakan *tools Python*
4. Proporsi data latih dan uji adalah 80% dan 20%

### 1.3. Jenis dan Sumber Data

Jenis data yang digunakan pada penelitian ini adalah data sekunder berupa *history* data harian harga saham perusahaan PT Bank Rakyat Indonesia Persero Tbk. Data sekunder ini diperoleh melalui pihak perantara bukan pihak terkait langsung, data tersebut diunduh melalui situs data tersebut diunduh melalui situs *yahoo finance*. Data yang digunakan adalah data sebanyak 1999 baris data, periode 2012 sampai 2022. Adapun atribut data sebagai berikut:

Tabel 2 2 Atribut Dataset

Atribut	Keterangan
Date	Tanggal yang tercatat pada waktu harga saham terjadi
Open	Harga pembukaan saham
High	Harga tertinggi saham
Low	Harga terendah saham
Close	Harga penutupan saham terakhir dalam satu hari
Adj Close	Harga penutupan saham yang telah disesuaikan dengan <i>close</i> ketika terjadi aksi atau keputusan perusahaan seperti <i>dividen</i> dan <i>stock split</i>
Volume	Jumlah transaksi saham yang terjadi dalam satu hari

### 1.4. Saham

Saham (*stock*) merupakan tanda pemilikan seseorang dalam suatu perusahaan atau perseroan yang berwujud selembar kertas, saham bersifat fluktuatif secara tidak menentu (Harris, 2003). Ada beberapa kondisi yang menyebabkan harga saham menjadi fluktuatif antara lain, kinerja dan kondisi perusahaan, kebijakan pemerintah, (Sandi, 2020).sedangkan menurut penelitian (Harsi Romli, 2017) Inflasi,nilai tukar uang,volume perdagangan, dan suku bunga SBI berpengaruh

terhadap volatilitas harga saham. Meskipun banyak faktor terhadap volatilitas harga saham. Para investor atau pemegang saham perlu menganalisa secara fundamental perusahaan sebagai bahan pertimbangan dalam melakukan investasi jangka pendek atau panjang. (Greis.Lilipaly,2014)

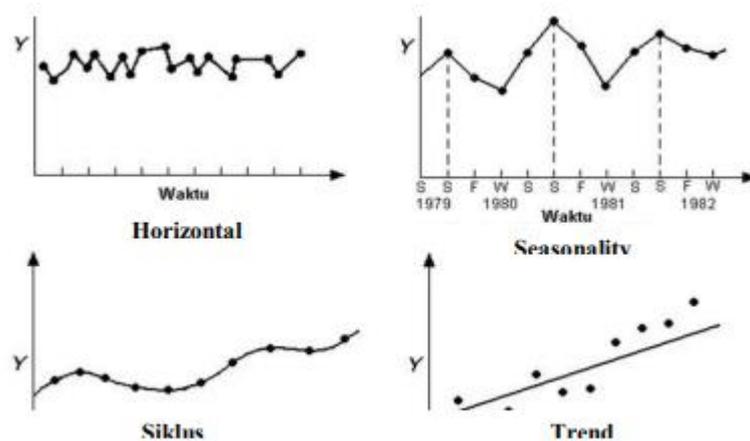
## 1.5. Peramalan

Peramalan merupakan proses memperkirakan berapa kebutuhan dimasa mendatang meliputi kuantitas,kualitas, waktu,dan lokasi dengan sistematis.(Zili, Derick Hendri and Kharis, 2022)

### 1.5.1. Jenis Pola Peramalan

Menurut Seto (2016) ada beberapa pola peramalan antara lain:

- 1) *Trend* (T) merupakan pola yang terjadi apabila ada kenaikan atau penurunan dari data secara gradual dalam kurun waktu panjang
- 2) *Seasonality* (S) merupakan pola musiman terjadi bila pola data berulang pada suatu periode tertentu: hari, mingguan, bulanan, triwulan dan,tahunan
- 3) *Horizontal* (H) atau Stationer, merupakan pola terjadi apabila nilai data berfluktuasi disekitar nilai rata – rata.
- 4) *Cycles* (C) adalah siklus suatu pola data yang terjadi setiap beberapa tahun biasanya dipengaruhi fluktuasi ekonomi jangka panjang.



Gambar 2. 1 Pola data

### 1.6. Metode ARIMA (*Autoregressive Integrated Moving Average*)

Metode ARIMA merupakan metode yang dibagi dalam 3 unsur yaitu: model AR (*Autoregressive*) yang merupakan model linear regresi berdasarkan lag dari data prediktor ( $p$ ), I (*Integrated*) adalah menghitung variasi pada waktu yang berbeda atau berapa kali dilakukan differencing untuk membuat suatu data stasioner ( $d$ ), dan MA (*Moving Average*) adalah melakukan rata – rata berjalan terhadap data runtun waktu tersebut ( $q$ ) (Wijesinghe and Rathnayaka, 2020) ARIMA biasa disebut juga sebagai metode Box-Jenkins yang merupakan metode dikembangkan oleh George Box dan Gwilym Jenkins pada tahun 1970 secara intensif. (Irawan, 2006). Model ARIMA ( $p,d,q$ ) bentuk matematis sebagai berikut.

$$Z_t = (1 + \phi_1)Z_{t-1} + (\phi_1 - \phi_2)Z_{t-2} + \dots + (\phi_p - \phi_{p-1})Z_{t-p} - \phi_p Z_{t-p-1} \\ + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

Dimana:

$p$	= Orde AR
$d$	= Orde <i>differencing</i>
$q$	= Orde MA
$Z_t$	= Nilai deret waktu Z pada waktu $t$
$a_t$	= Nilai dari proses <i>white noise</i> pada waktu $t$
$a_{t-1} \dots a_{t-q}$	= Nilai dari proses <i>white noise</i> pada masing masing selang waktu $t-1, t-2 \dots, t-q$ , nilai $a$ sebagai variabel bebas
$Z_{t-1} \dots Z_{t-p}$	= Nilai deret waktu Z pada masing masing selang $t-1, t-2, \dots, t-p$
$\phi$	= Parameter autoregressi (AR)
$\theta$	= Parameter moving average (MA)

#### 1.6.1. Stasioneritas

Suatu data *time series* dapat dikatakan stasioner jika data tidak memiliki trend ataupun seasonal (musiman) pola tertentu pada suatu runtun waktu. Untuk data yang tidak stasioner dibagi menjadi tiga jenis yaitu data yang tidak stasioner dalam rata – rata (mean), data tidak stasioner dalam varians dan data yang tidak stasioner dalam rata – rata (mean) dan varians. Tahap untuk menstasionerkan data juga ada beberapa macam, seperti untuk data yang tidak stasioner dalam rata – rata (mean)

harus melakukan pembedaan atau differencing, sedangkan untuk data yang tidak stasioner dalam varians maka data harus ditransformasi. (Marvillia, 2017).

### 1.6.2. Augmented Dickey-Fuller (ADF)

Salah satu konsep formal yang dipakai untuk mengetahui stasioneritas data adalah melalui uji akar unit (unit root test). Uji ini merupakan pengujian yang populer, dikembangkan oleh David Dickey dan Wayne Fuller dengan sebutan Augmented Dickey-Fuller (ADF) Test. Persamaan matematis sebagai berikut:

$$\Delta Z_{t-1} = \lambda Z_{t-1} + \ell_t$$

dimana:

$\lambda$  = nilai koefisien untuk nilai pengamatan pada waktu ke t-1

$\Delta Z_{t-1}$  = Selisih nilai antara data ke -t dengan data ke t-1

$\ell_t$  = Variabel residual

$Z_t$  = nilai pengamatan pada waktu ke t

$Z_{t-1}$  = nilai pengamatan pada waktu ke t-1

Hipotesis yang digunakan pada pengujian *Augmented Dickey Fuller* adalah:

$H_0 : \delta = 0$  (data tidak stasioner)

$H_1 : \delta = 1$  (data stasioner)

### 1.6.3. Diferensiasi

Diferensiasi merupakan mengubah deret waktu nonstasioner menjadi stasioner dengan menghasilkan deret baris baru yang didapatkan dari perbedaan nilai variabel dalam satu periode dengan periode sebelumnya.

$$Y'_t = Y_t - Y_{t-1}$$

Dimana:

$Y'_t$  = data yang akan di differencing

$Y_t$  = merupakan data waktu tertentu

t = index waktu

#### 1.6.4. *Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF)*

Autokorelasi (ACF) merupakan korelasi dari sebuah data time series untuk selang waktu (lag) yang berlainan. Autokorelasi dapat digunakan untuk menentukan ada tidaknya faktor musiman (seasonality). Selain itu, autokorelasi dapat digunakan untuk menentukan kestasioneran suatu data. Autokorelasi parsial (PACF) merupakan pengembangan dari autokorelasi, yaitu dengan cara menghilangkan dependensi linear pada variabel  $Z_{t+1}$ ,  $Z_{t+2}$ , ..., dan  $Z_{t+k-1}$ , sehingga diperoleh bentuk korelasi baru (Wei, 2016).

Menurut Gujarati 2003 Pola ACF dan PACF sebagai berikut:

Tabel 2 3 Pola ACF dan PACF

ACF	PACF
AR(p) Turun cepat secara eksponensial	Ma(q) Turun cepat secara eksponensial

#### 1.6.5. *Model Autoregressive (AR)*

Pada tahun 1926, Yule memperkenalkan metode *autoregressive*. dilanjutkan pada tahun 1931 metode AR dikembangkan oleh Walker. Metode AR diasumsikan bahwa data saat ini dipengaruhi oleh data lampau atau sebelumnya. Model Autoregressive dengan ordo p disingkat menjadi AR (p) atau ARIMA (p,0,0)

Bentuk persamaan dari model AR (*Autoregressive*) dinyatakan sebagai berikut:

$$X_t = \mu' + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t$$

Dimana:

$\mu'$  = suatu konstanta

$X_t$  = data time series pada waktu ke t

$\phi_p$  = parameter autoregresif ke-p

$e_t$  = nilai kesalahan pada saat t

### 1.6.6. Model *Moving Average* (MA)

Pada tahun 1973 MA (*Moving Average*) pertama kali diperkenalkan oleh Slutsky, dengan ordo  $q$  atau ARIMA (0,0, $q$ ). Menurut Subagyo pada tahun 2008 model (*Moving Average*) atau rata-rata bergerak merupakan peramalan yang dilakukan dengan mengambil sekelompok nilai pengamatan, mencari nilai rata-rata tersebut sebagai ramalan untuk periode yang akan datang.

Bentuk persamaan dari model MA (*Moving Average*) dinyatakan sebagai berikut:

$$X_t = \mu' + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_k e_{t-k}$$

Dimana:

$\mu'$  = suatu konstan .

$\theta_1$  sampai  $\theta_k$  adalah parameter-parameter moving average

$e_{t-k}$  = nilai kesalahan pada saat  $t-k$ .

### 1.7. Metode *Exponential Smoothing*

Pada tahun 1950 pertama kali Metode (*exponential smoothing*) atau model peramalan pemulusan eksponensial diperkenalkan oleh Robert G. Brown melalui karya asli (1959, 1969). Pemulusan eksponensial (*exponential smoothing*) merupakan metode peramalan yang digunakan untuk meramalkan masa yang akan datang dengan melakukan proses pemulusan (*smoothing*) dengan menghasilkan data ramalan yang lebih kecil nilai kesalahannya. Dalam pemulusan (*smoothing*) eksponensial terdapat satu atau lebih parameter pemulusan yang ditentukan secara eksplisit dan hasil pilihan menentukan bobot yang dikenakan pada nilai observasi (Makridakis, 1999). Dalam penelitian ini menggunakan *Single Exponential Smoothing* (SES). Metode *Single Exponential Smoothing* adalah model peramalan dimana data diberi bobot oleh fungsi exponential dari rata-rata bergerak (Zainurrosid, Wardana and Siregar, 2022). Metode ini hanya mempertimbangkan satu level komponen, yaitu level aktual atau rata-rata data observasi masa lalu. SES cocok untuk data yang stasioner atau hampir stasioner. Dalam SES, peramalan

dihitung dengan menggabungkan data observasi terakhir dengan peramalan sebelumnya menggunakan faktor perataan (*smoothing factor*). bentuk matematis model sebagai berikut:

$$F_t = F_{(t-1)} + \alpha (A_{(t-1)} - F_{(t-1)})$$

Dimana:

$F_t$  : Nilai peramalan untuk periode waktu t

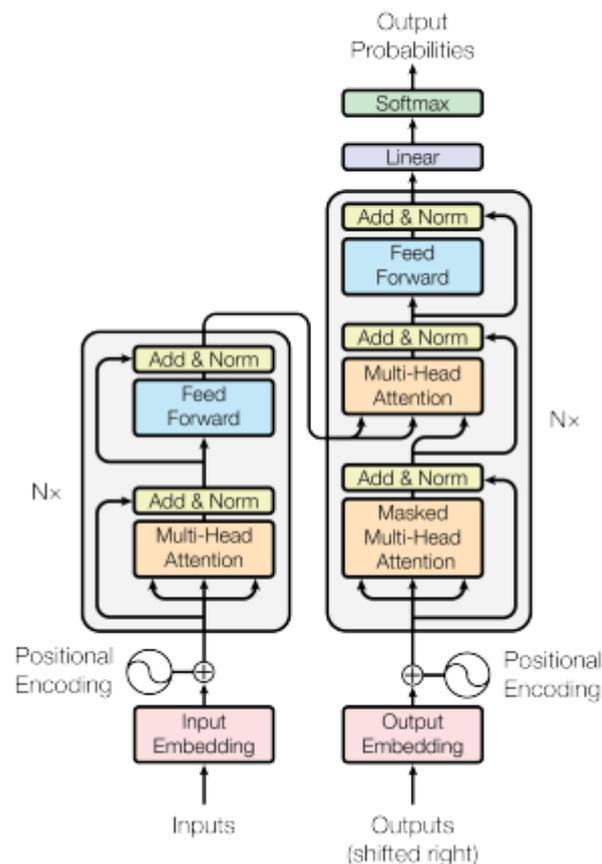
$F_{(t-1)}$  : Nilai peramalan untuk 1 periode waktu sebelum t

$\alpha$  : Nilai Konstanta *smoothing*

$A_{(t-1)}$  : Nilai aktual untuk 1 periode waktu ke t

## 1.8. Transformers

Model Transformers merupakan model transduksi pertama yang mengandalkan sepenuhnya pada perhatian model untuk menghitung representasi input dan outputnya tanpa menggunakan RNN atau konvolusi yang disejajarkan secara berurutan., model transformer memiliki struktur encoder – decoder (Vaswani, A. et al. 2017). arsitektur Transformer. Transformer adalah arsitektur jaringan saraf yang menggunakan mekanisme perhatian (*attention mechanism*) untuk memproses urutan data Penerapan awal transformers adalah pada model Bahasa Alami seperti Google's BERT dan OpenAI's GPT-2, yang mengubah industri pemrosesan bahasa alami dan menghasilkan hasil yang jauh lebih baik daripada model sebelumnya Arsitektur metode transformer , Transformers memperoleh popularitas besar dalam bidang pemrosesan bahasa alami (*natural language processing/NLP*) dan berbagai tugas terkait seperti pemodelan bahasa, penerjemahan mesin, dan pemahaman teks. Produk implementasi Transformer saat ini terkenal adalah ChatGPT. Dalam konteks ChatGPT, Transformer digunakan untuk memahami konteks dan menghasilkan respon yang sesuai berdasarkan masukan pengguna sebelumnya. Dengan memperhatikan seluruh konteks dari dialog, Transformer dapat menangkap hubungan dan dependensi antara kata-kata dalam teks yang relevan untuk menghasilkan respon yang lebih informatif dan koheren. Arsitektur Transformer ditunjukkan pada gambar dibawah ini.



Gambar 2. 2 Arsitektur Transformers

Pada gambar di atas merupakan arsitektur dari model *Transformers*, terdiri dari dua bagian yaitu *Encoder* dan *Decoder*,

### 1.8.1. Encoder

Encoder adalah bagian dari model yang bertugas mengolah informasi masukan (input sequence) dan menghasilkan representasi numerik dari informasi tersebut dalam bentuk vektor. Encoder pada dasarnya terdiri dari beberapa lapisan yang masing-masing terdiri dari dua jenis sub-layer, yaitu *multi-head self-attention layer* dan *position-wise feedforward layer*. Sub-layer *multi-head self-attention* berguna untuk mempelajari hubungan antar kata pada masukan, sedangkan sub-layer *position-wise feedforward* bertugas untuk memperkuat representasi numerik tersebut. Menurut Vaswani et al. (2017) - Encoder Transformer menggunakan blok transformasi bertingkat (*layered transformation blocks*) yang terdiri dari mekanisme perhatian yang dilatih untuk mengubah urutan input menjadi representasi kontekstual yang lebih tinggi.

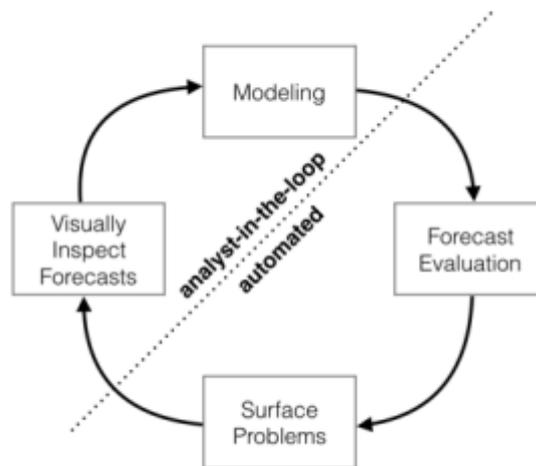
### 1.8.2. Decoder

Decoder adalah bagian dari model yang bertugas menghasilkan keluaran (output sequence) berdasarkan representasi numerik dari informasi masukan yang dihasilkan oleh encoder. Seperti halnya dengan encoder, decoder juga terdiri dari beberapa lapisan, namun memiliki satu jenis sub-layer tambahan, yaitu *masked multi-head self-attention layer*. Sub-layer ini berguna untuk memperhatikan posisi-posisi sebelumnya pada keluaran dan memastikan bahwa decoder hanya dapat "melihat" informasi yang seharusnya telah tersedia pada saat ini. Kedua komponen tersebut dihubungkan oleh sebuah mekanisme attention, dimana output dari encoder digunakan untuk memberikan "perhatian" pada beberapa bagian tertentu dari informasi masukan yang relevan dengan prediksi yang akan dilakukan oleh decoder.

### 1.9. FB - Prophet

Prophet adalah paket sumber terbuka (untuk Python dan R) yang dirilis oleh tim inti Ilmu Data (*Data Science*) Facebook untuk memperkirakan data deret waktu berdasarkan model adiktif di mana tren non-linier sesuai dengan musiman tahunan, mingguan, dan harian, ditambah efek liburan (Garlapati *et al.*, 2021). Prophet bekerja paling baik dengan deret waktu yang memiliki efek musiman yang kuat dan beberapa musim data historis. Prophet juga sangat bagus untuk data yang hilang dan pergeseran tren, dan biasanya menangani data hilang (*outlier*) dengan baik (Menculini *et al.*, 2021) keunggulan Facebook Prophet sebagai berikut.

1. Kecepatan Komputasi Prophet menggunakan metode perhitungan yang dioptimalkan untuk meningkatkan kecepatan komputasi. Pustaka ini mampu menghasilkan peramalan dengan cepat, bahkan pada dataset yang besar, sehingga berguna dalam situasi yang membutuhkan peramalan secara real-time.
2. Kemampuan Menangani Fitur-fitur Khas Data Waktu: Prophet menawarkan kemampuan yang kuat dalam menangani fitur-fitur khas data waktu, seperti tren musiman dan efek hari libur. Ini memungkinkan untuk memperhitungkan dan mengintegrasikan fitur-fitur ini ke dalam model peramalan dengan mudah dan akurat



Gambar 2. 3 Skema Model Facebook Prophet

Prophet menggunakan *Decomposable Time Series Model* dengan tiga komponen utama yaitu *Seasonal*, *Trends*, dan *Holiday*. persamaan matematis metode Prophet sebagai berikut:

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

Dengan:

$g(t)$  = Trend

$s(t)$  = Periodik atau *Seasonality*

$h(t)$  = Efek hari libur atau *Holiday*

$\epsilon_t$  = Error yang tidak diakomodasi oleh model

Menurut Taylor dan Letham pada tahun 2017, Prophet memiliki keunggulan yaitu:

- 1) Prophet tidak perlu menangani data yang bernilai kosong (*Missing Value*)
- 2) Dapat dengan mudah mengakomodasi data musiman.
- 3) Proses *Training Model* atau pelatihan sangat cepat.

### 1.10. Min-max normalization

Min-max normalisasi merupakan salah satu teknik normalisasi data yang dapat diterapkan pada data waktu stasioner dengan mengubah skala antara rentang 0 atau 1. (Munkhdalai *et al.*, 2019). Persamaan matematis sebagai berikut.

$$u_i = \frac{v - \min_a}{\max_a - \min_a}$$

Dimana:

$u_i$  : Data sudah di normalisasi

$v$  : Data belum di normalisasi

$\max_a$  : Nilai maksimum

$\min_a$  : Nilai minimum

### 1.11. Mean Absolute Percentage Error (MAPE)

MAPE (*Mean Absolute Percentage Error*) digunakan untuk menghitung rata – rata persentase antara data saham aktual dengan harga saham telah di ramalkan.

Dengan persamaan matematis sebagai berikut:

$$MAPE = \sum_{i=1}^n \left| \frac{x - y_i}{x} \right| \times 100\%$$

Dimana:

$n$  = jumlah data saham

$x$  = data aktual saham

$y_i$  = data prediksi harga saham

Tabel 2 4 Nilai MAPE

Nilai MAPE	Interpretasi
$\leq 10\%$	Hasil prediksi sangat baik
10% - 20%	Hasil Prediksi baik
20% - 50%	Hasil prediksi wajar (layak)
$>50\%$	Hasil prediksi tidak akurat