

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1. Permasalahan**

Industri asuransi memiliki peran yang sangat vital dalam perekonomian, terutama dalam memberikan perlindungan finansial terhadap risiko yang tidak terduga. Salah satu produk asuransi yang memiliki dampak signifikan adalah asuransi kesehatan, khususnya asuransi penyakit kritis (*critical illness*). Produk ini memberikan perlindungan kepada tertanggung dari biaya besar yang timbul akibat penyakit berat seperti kanker, stroke, dan penyakit jantung. Berbagai perusahaan asuransi sudah menawarkan produk asuransi kesehatan penyakit kritis untuk memberikan perlindungan maksimal kepada nasabahnya. PT X merupakan salah satu perusahaan asuransi yang memiliki produk asuransi kesehatan penyakit dalam yang menjadi ceding atau nasabah dari PT Reasuransi Indonesia Utama. Sebagai perusahaan reasuransi, PT Reasuransi Indonesia Utama harus menentukan besar premi yang harus dibayar nasabah Asuransi X agar tidak mengalami kerugian saat proses klaim. Dalam menentukan besar premi, baik perusahaan asuransi maupun reasuransi harus memperhitungkan banyak hal, salah satunya data rekam jejak klaim yang sudah terjadi sebelumnya.

Peramalan klaim asuransi kesehatan bukanlah tugas yang mudah. Besar klaim dipengaruhi oleh berbagai faktor, termasuk tren kesehatan masyarakat, perubahan demografi, kebijakan kesehatan pemerintah, dan faktor ekonomi. Metode tradisional sering kali tidak mampu menangkap kompleksitas dan non-linearitas data klaim asuransi. Oleh karena itu, diperlukan pendekatan yang lebih canggih dan adaptif. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja dua model peramalan, yaitu *Autoregressive Integrated Moving Average* (ARIMA) dan *Long Short-Term Memory* (LSTM), dalam memprediksi besar klaim produk asuransi kesehatan *critical illness* milik PT X. Dengan menggunakan kedua model ini, diharapkan dapat diperoleh model

peramalan yang lebih akurat dan handal, sehingga dapat membantu perusahaan dalam pengambilan keputusan yang lebih baik.

## **4.2. Metodologi Penelitian**

Metodologi penelitian merupakan penjabaran dari proses penelitian secara keseluruhan. Dalam metodologi penelitian ini akan diperkenalkan sumber data, variabel penelitian, serta desain dan metode penelitian.

### **4.2.1. Sumber Data**

Data yang akan digunakan dalam penelitian ini adalah data historis besar klaim produk asuransi kesehatan penyakit kritis PT X dalam rentang waktu Januari 2019 – November 2023.

### **4.2.2. Variabel Penelitian**

Penelitian ini akan meramalkan besar klaim produk asuransi kesehatan penyakit kritis milik PT X bernama A dan B sebanyak 59 data yang terdiri dari data bulanan sejak Januari 2019 hingga November 2023.

### **4.2.3. Desain dan Metode Penelitian**

Penelitian ini dilakukan dengan desain penelitian terapan menggunakan pendekatan kuantitatif. Proses penelitian dimulai dengan memahami cara perhitungan premi untuk produk asuransi kesehatan penyakit kritis milik PT X, yaitu A dan B, berdasarkan informasi dari aktuaris perusahaan. Setelah itu dilakukan pengumpulan data klaim dari sumber yang telah ditetapkan perusahaan dan menganalisis data tersebut sesuai dengan permasalahan penelitian. Analisis data yang dilakukan adalah analisis peramalan (*forecasting*) dengan membandingkan dua metode, yaitu *Autoregressive Integrated Moving Average* (ARIMA) dan *Long Short-Term Memory* (LSTM).

### 4.3. Data *Pre-Processing*

Proses analisis pada penelitian ini diawali dengan pra pemrosesan data runtun waktu bulanan besar klaim produk asuransi kesehatan A dan B milik PT X dalam rentang waktu Januari 2019 hingga November 2023 sebanyak 23 titik data. *Pre-processing* data akan dilakukan dengan melakukan pengecekan terhadap *missing value* dan analisis deskriptif data.

#### 4.3.1. Pengecekan *Missing Value*

Sebelum memasuki tahapan analisis, akan dilakukan pengecekan terhadap missing value pada data besar klaim terlebih dahulu.

```
Jumlah missing value
Claim      0
Period     0
```

**Gambar 4.1. Pengecekan *Missing Value* Data Besar Klaim**



**Gambar 4.2. Plot Representasi Availabilitas *Missing Value***

Melalui gambar 4.1 dan 4.2, terlihat bahwa dalam data besar klaim tidak terdapat *missing value*. Pengecekan *missing value* dilakukan secara deskriptif dengan menggunakan *software Python*. Oleh karena itu, data besar klaim dapat dilanjutkan ke tahap analisis.

#### 4.3.2. Analisis Deskriptif Data

Analisis deskriptif data dilakukan untuk mengetahui karakteristik dari data yang digunakan. Dalam menganalisis data secara deskriptif, data besar klaim akan dipisah berdasarkan tahun.

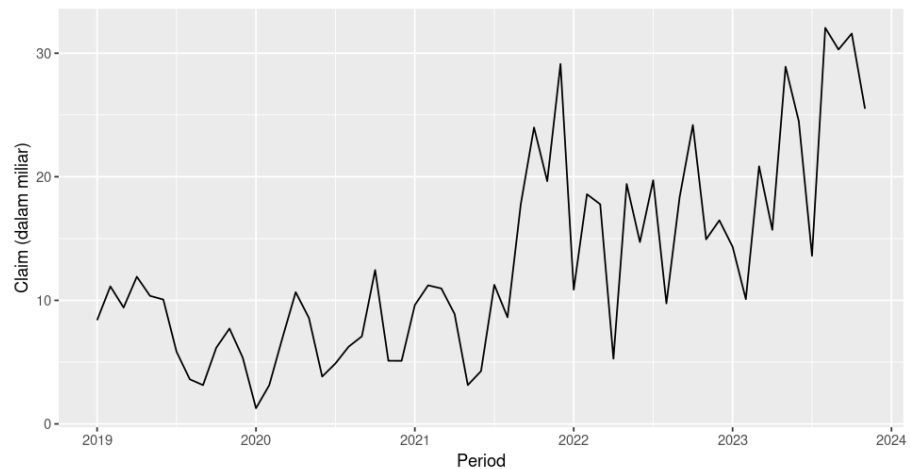
Tabel 4.1. Analisis Deskriptif Data Besar Klaim (dalam miliar rupiah)

<b>Periode</b>	<b>Mean</b>	<b>Min</b>	<b>Max</b>	<b>Standar Deviasi</b>
2019	7,759	3,14	11,91	2,929
2020	6,277	1,27	12,45	3,152
2021	13,21	3,143	29,123	7,851
2022	15,837	5,28	24,184	5,142
2023	22,493	10,099	32,048	7,982

Tabel 4.1 menunjukkan karakteristik data besar klaim dalam miliar rupiah. Berdasarkan tabel tersebut, terlihat bahwa besar klaim produk asuransi kesehatan dalam rentang waktu 2019 hingga 2023 sangat bervariasi. Pada tahun 2019, rata-rata besar klaim produk asuransi kesehatan adalah sebesar Rp 7.759.000.000,00 dengan standar deviasi Rp 2.929.000.000,00. Pada tahun 2020, rata-rata besar klaim produk asuransi kesehatan adalah sebesar Rp 6.227.000.000,00 dengan standar deviasi Rp 3.152.000.000,00. Pada tahun 2021, rata-rata besar klaim produk asuransi kesehatan adalah sebesar Rp 13.210.000.000,00 dengan standar deviasi Rp 7.851.000.000,00. Pada tahun 2022, rata-rata besar klaim produk asuransi kesehatan adalah sebesar Rp 15.837.000.000,00 dengan standar deviasi Rp 5.142.000.000,00. Pada tahun 2023, rata-rata besar klaim produk asuransi kesehatan adalah sebesar Rp 22.493.000.000,00 dengan standar deviasi Rp 7.982.000.000,00.

#### 4.3.3. Eksplorasi Data

Setelah melalui tahap analisis deskriptif data, akan ditinjau lebih jauh visualisasi dari data besar klaim.



**Gambar 4.3. Time Series Plot Data Besar Klaim**

Pada gambar 4.3, terlihat bahwa data besar klaim cenderung fluktuatif karena mengalami kenaikan yang cukup signifikan selama rentang waktu tahun 2019 hingga 2023. Selama tahun 2019, besar klaim berada pada kisaran Rp 1.000.000.000,00 hingga Rp 10.000.000.000,00. Namun pada tahun 2023, besar klaim sudah berada pada kisaran Rp 20.000.000.000,00 hingga Rp 30.000.000.000,00. Tidak ada pola musiman pada data besar klaim, maka dari itu analisis peramalan tidak akan menggunakan *Seasonal Autoregressive Integrated Moving Average* (SARIMA), melainkan *Autoregressive Integrated Moving Average* (ARIMA).

#### **4.4. Peramalan dengan *Autoregressive Integrated Moving Average* (ARIMA)**

Dalam melakukan *forecasting* dengan metode ARIMA, data harus melalui beberapa tahapan sebelum memperoleh hasil peramalan. Data yang digunakan dalam ARIMA merupakan data yang stasioner, oleh karena itu harus dilakukan uji stasioneritas data terlebih dahulu. Apabila data belum stasioner, akan dilakukan transformasi data hingga data memenuhi asumsi stasioneritas. Selanjutnya, akan ditentukan orde  $p$  dan  $q$  melalui plot *autocorrelation* (ACF) dan *partial autocorrelation* (PACF), orde  $d$  diperoleh melalui banyak

diferensiasi saat transformasi data. Peramalan dengan metode ARIMA dilakukan dengan menggunakan orde p, q, dan d yang dimiliki,

#### 4.4.1. Stasioneritas Data

Tahap awal dari metode ARIMA adalah menguji stasioneritas data yang dimiliki. Uji ini dilakukan dengan tujuan meminimalkan bias pada model hasil peramalan. Uji stasioneritas data dilakukan menggunakan uji *Augmented Dickey Fuller* (ADF) dengan *software R Studio*.

```
Augmented Dickey-Fuller Test  
data: data  
Dickey-Fuller = -3.5112, Lag order = 3, p-value = 0.04838  
alternative hypothesis: stationary
```

**Gambar 4.4. Hasil *Augmented Dickey-Fuller Test* Data Besar Klaim**

Berdasarkan gambar 4.4, diperoleh hasil uji ADF pada data besar klaim memiliki P - Value sebesar 0,04838 yang lebih kecil dari nilai tingkat signifikansi ( $\alpha$ ) sebesar 0,05 atau 5%. Maka dapat disimpulkan bahwa data besar klaim sudah stasioner dengan nilai mutlak Dickey-Fuller sebesar 3,5112. Meskipun data sudah stasioner, akan dilakukan pengecekan orde diferensiasi yang paling tepat untuk proses transformasi data.

```
Warning message in ndiffs(log(data), "adf"):  
"Specified alpha value is larger than the maximum, setting alpha=0.1"  
1
```

**Gambar 4.5. Orde Diferensiasi untuk Data Besar Klaim**

Orde diferensiasi (d) yang paling agar data yang dimiliki bersifat stasioner berdasarkan gambar 4.5 adalah 1. Maka dari itu, akan dilakukan diferensiasi terhadap data dengan transformasi dan transformasi logaritma 1.

#### 4.4.2. Transformasi Data

Transformasi data besar klaim akan dilakukan dengan orde diferensiasi 1.

```
Augmented Dickey-Fuller Test
data: ddif1
Dickey-Fuller = -6.2919, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

**Gambar 4.6. Hasil Uji ADF untuk Transformasi Diferensi 1**

```
Augmented Dickey-Fuller Test
data: dtrans1
Dickey-Fuller = -7.4198, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

**Gambar 4.7. Hasil Uji ADF untuk Transformasi Diferensi Log1**

Dari *Augmented Dickey-Fuller Test* yang telah dilakukan, diperoleh tabel hasil uji stasioneritas data sebagai berikut

**Tabel 4.2. Tabel Hasil Transformasi Data**

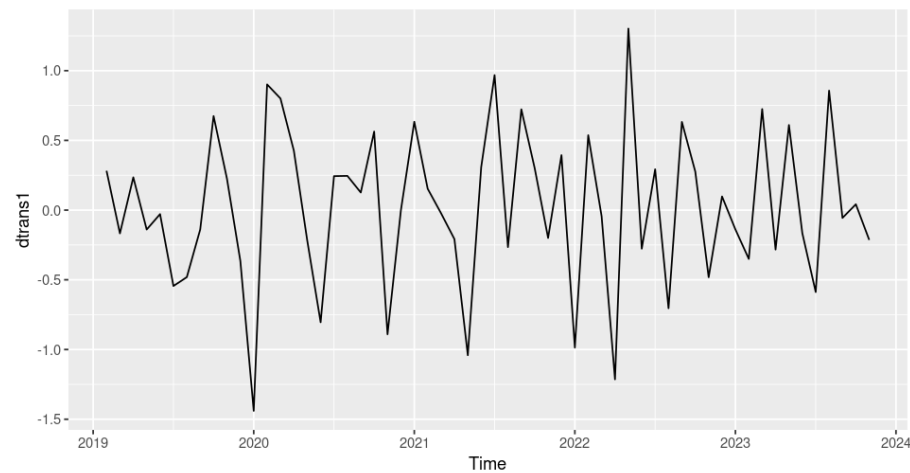
	Dickey-Fuller	P – Value
Data	3,5112	0,04838
Data setelah transformasi diferensi 1	6,2919	0,01
Data setelah transformasi diferensi log1	7,4198	0,01

Hasil transformasi data yang paling baik, dapat ditentukan menggunakan nilai mutlak dari Dickey-Fuller dan P – Value. Data yang paling stasioner memiliki nilai mutlak Dickey-Fuller terbesar dengan P – Value bernilai lebih kecil dari tingkat signifikansi yang telah ditentukan. Berdasarkan hasil transformasi yang dimiliki, dapat disimpulkan bahwa data setelah melalui transformasi diferensi log1 merupakan hasil transformasi terbaik karena memiliki nilai mutlak terbesar yaitu 7,4198. Selain itu, diperoleh nilai P – Value sebesar 0,01 yang lebih kecil dari nilai  $\alpha$  sebesar 0,05. Maka, data yang akan digunakan untuk analisis

selanjutnya merupakan data hasil transformasi diferensi log1 dan orde d dari model ARIMA adalah 1.

#### 4.4.3. Identifikasi Model

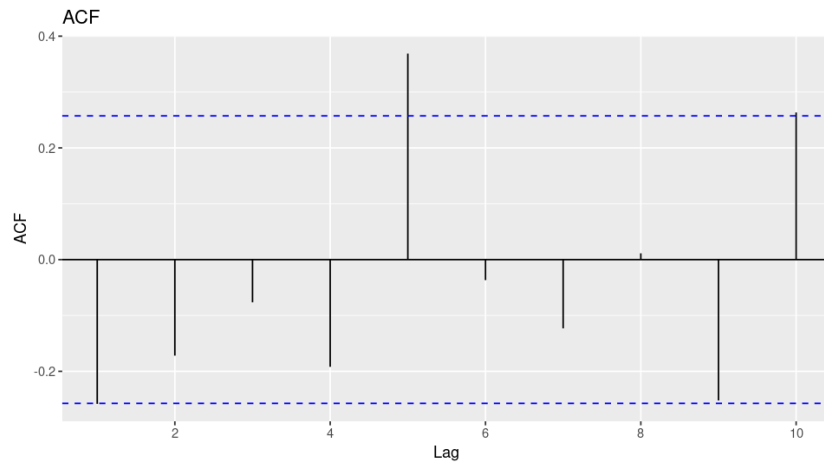
Setelah data train melalui tahap transformasi, diperoleh plot data sebagai berikut



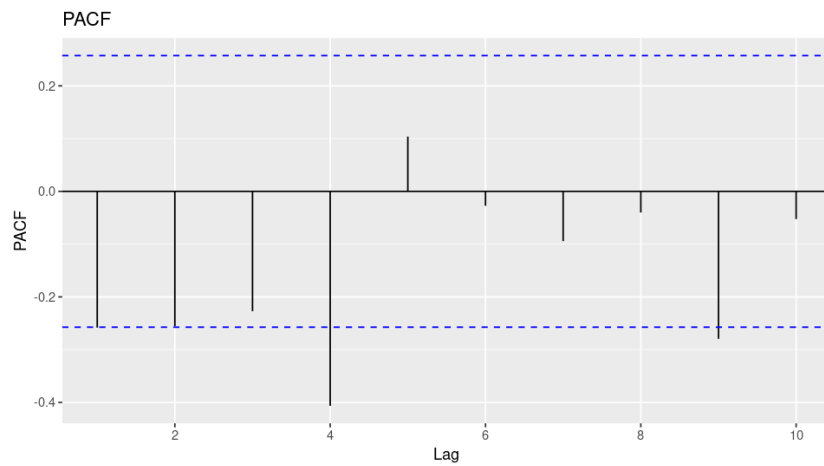
**Gambar 4.8. Plot Data Besar Klaim setelah Transformasi dengan Diferensi Log1**

Terlihat bahwa data sudah cukup stasioner karena memiliki sifat statistika seperti mean, varians, dan autokorelasi dari data yang tetap konstan sepanjang periode waktu. Selanjutnya, akan dilakukan identifikasi model untuk menentukan orde  $p$  dan  $q$  dari model ARIMA. Pengidentifikasian model dilakukan dengan mencermati plot *autocorrelation function* (ACF) dan plot *partial autocorrelation function* (PACF). Plot ACF digunakan untuk mengidentifikasi orde *Moving Average* yaitu  $q$ , sedangkan plot PACF digunakan untuk mengidentifikasi orde *Autoregressive* yaitu  $p$ .





**Gambar 4.9. Plot *Autocorrelation Function* Data Hasil Transformasi**



**Gambar 4.10. Plot *Partial Autocorrelation Function* Data Hasil Transformasi**

Dari gambar, dapat terlihat bahwa fungsi ACF signifikan pada lag ke-1, sedangkan fungsi PACF signifikan pada lag ke-4. Oleh karena itu, dapat disimpulkan bahwa orde MA (q) bernilai 1 dan orde AR (p) bernilai 4.

#### **4.4.4. Pembentukan Model**

Setelah melalui tahap transformasi dan identifikasi model, diperoleh model ARIMA(4,1,1) dengan orde p yaitu 4, orde d yaitu 1, dan orde q yaitu 1. Dengan orde tersebut, akan dilakukan estimasi parameter model ARIMA(4,1,1) dan turunan modelnya. Turunan model

atau underfitting dari ARIMA(4,1,1) adalah ARIMA(4,1,0), ARIMA(3,1,1), ARIMA(3,1,0), ARIMA(2,1,1), ARIMA(2,1,0), ARIMA(1,1,1), ARIMA(0,1,1), dan ARIMA(1,1,0). Berikut adalah hasil estimasi parameter setiap modelnya.

**Tabel 4.3. Tabel Ringkasan Estimasi Parameter Model  
ARIMA(4,1,1) dan Turunan Modelnya**

p,d,q	P - Value					
	Constant	AR(1)	AR(2)	AR(3)	AR(4)	MA(1)
4,1,1	0,481	0,004	0,000	0,002	0,000	0,338
4,1,1	-	0,004	0,000	0,003	0,000	0,315
4,1,0	0,452	0,000	0,000	0,002	0,001	-
4,1,0	-	0,000	0,000	0,003	0,001	-
3,1,1	0,000	0,001	0,682	0,994	-	0,000
3,1,1	-	0,463	0,326	0,450	-	0,005
3,1,0	0,586	0,003	0,012	0,079	-	-
3,1,0	-	0,003	0,014	0,085	-	-
2,1,1	0,000	0,001	0,665	-	-	0,000
2,1,1	-	0,095	0,373	-	-	0,000
2,1,0	0,663	0,047	0,011	-	-	-
2,1,0	-	0,012	0,049	-	-	-
1,1,1	0,000	0,001	-	-	-	0,000
1,1,1	-	0,052	-	-	-	0,000
0,1,1	0,371	-	-	-	-	0,000
0,1,1	-	-	-	-	-	0,000
1,1,0	0,740	0,043	-	-	-	-
1,1,0	-	0,044	-	-	-	-

Model yang signifikan ditentukan dengan mencermati nilai P – Value dari masing-masing parameternya. Parameter yang signifikan memiliki nilai P – Value lebih kecil dari tingkat signifikansi atau  $\alpha$  yaitu 0,05. Berdasarkan tabel 4.3, dapat disimpulkan bahwa model yang signifikan adalah ARIMA(4,1,0), ARIMA(2,1,0), ARIMA(0,1,1), dan ARIMA(1,1,0) tanpa konstanta, serta ARIMA(1,1,1) dengan konstanta.

Selanjutnya akan dilakukan pemilihan model terbaik dengan menggunakan *diagnostic checking* dan akurasi tiap model.

#### 4.4.5. *Diagnostic Checking* dan Pemilihan Model Terbaik

*Diagnostic checking* dalam penelitian ini terdiri dari asumsi homoskedastisitas, normalitas, no autokorelasi, dan residual. Model yang memenuhi keempat asumsi diagnostik tersebut akan ditinjau lebih lanjut nilai akurasi hingga diperoleh model ARIMA yang paling tepat. Berikut adalah hasil dari *diagnostic checking* tiap model.

**Tabel 4.4. Hasil *Diagnostic Checking* Model yang Signifikan**

p,d,q	Homoskedastisitas	Normalitas	No Autokorelasi	Residual
4,1,0	Terpenuhi	Terpenuhi	Terpenuhi	Terpenuhi
2,1,0	Terpenuhi	Terpenuhi	Terpenuhi	Tidak terpenuhi
1,1,1	Terpenuhi	Tidak terpenuhi	Terpenuhi	Tidak terpenuhi
1,1,0	Terpenuhi	Terpenuhi	Terpenuhi	Tidak terpenuhi
0,1,1	Terpenuhi	Tidak terpenuhi	Terpenuhi	Tidak terpenuhi

Berdasarkan tabel 4.4, diperoleh model ARIMA(4,1,0) tanpa konstanta memenuhi keempat asumsi diagnostik. Maka dari itu, model ARIMA dengan orde p yaitu 4, orde d yaitu 1, dan orde q yaitu 0 merupakan model ARIMA terbaik untuk meramalkan data besar klaim.

#### 4.4.6. Peramalan Data *Train* dan Metrik Evaluasi

Setelah memperoleh ARIMA(4,1,0) tanpa konstanta sebagai model ARIMA terbaik, langkah selanjutnya adalah melakukan peramalan terhadap data besar klaim untuk 7 periode berikutnya dan memperoleh metrik evaluasi dari hasil peramalan tersebut.

Model	LogLik	AIC	BIC
<chr>	<dbl>	<dbl>	<dbl>
tc410	-39.05265	88.1053	98.40751

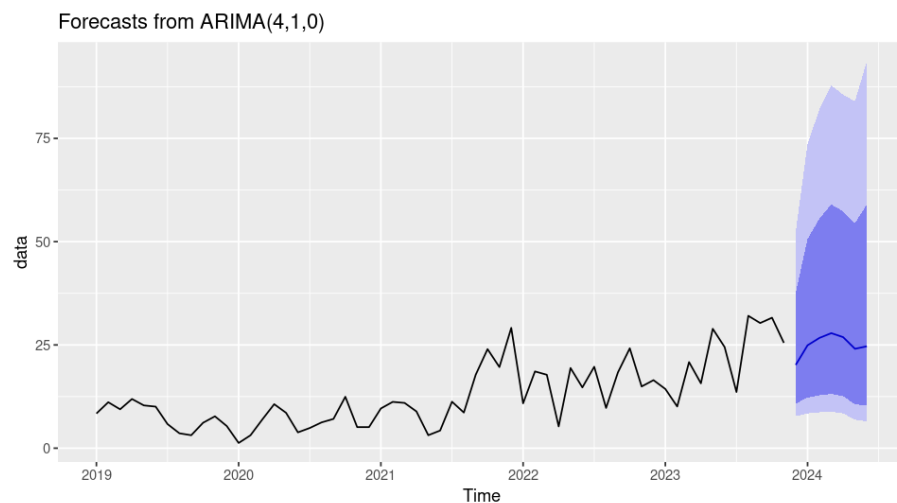
A matrix: 1 × 7 of type dbl

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	1.420499	5.654335	4.085936	-7.504518	39.03255	0.6046154	-0.02136124

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Dec 2023	20.15816	10.79349	37.64781	7.754438	52.40242
Jan 2024	24.91645	12.26277	50.62721	8.425530	73.68434
Feb 2024	26.69536	12.81771	55.59823	8.692414	81.98436
Mar 2024	27.85070	13.14570	59.00495	8.834505	87.79911
Apr 2024	26.87953	12.60496	57.31944	8.441948	85.58559
May 2024	24.06062	10.62735	54.47390	6.895442	83.95595
Jun 2024	24.66489	10.30723	59.02229	6.494490	93.67273

**Gambar 4.11. Hasil Peramalan dan Metrik Evaluasi dengan ARIMA(4,1,0) Tanpa Konstanta**

Berdasarkan hasil perhitungan akurasi dan peramalan data besar klaim, diperoleh 7 titik *forecast* untuk 7 periode ke depan, yakni bulan Desember 2023 hingga Juni 2024. Selain itu, diperoleh nilai metrik evaluasi MAE sebesar 4,086, RMSE sebesar 5,654 dan MAPE sebesar 39,033. Plot data hasil peramalan 7 bulan berikutnya adalah sebagai berikut.



**Gambar 4.12. Plot Hasil Peramalan dengan Model ARIMA(4,1,0) Tanpa Konstanta untuk 7 Periode Berikutnya**

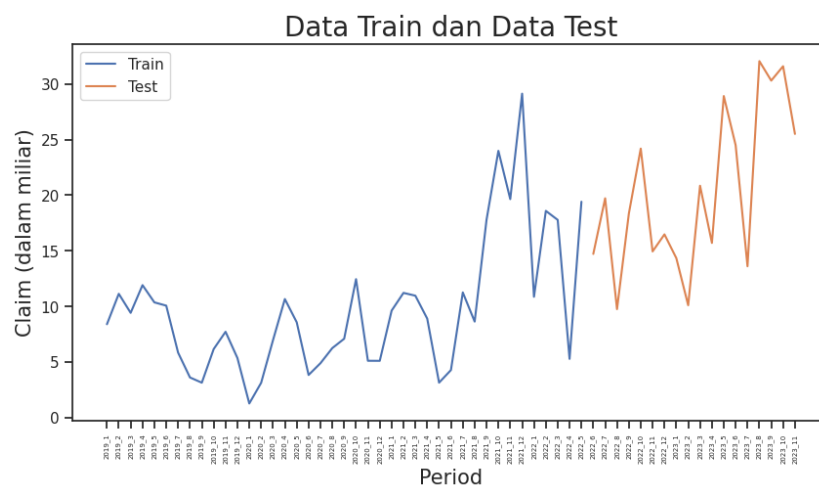
Melalui gambar, terlihat bahwa besar klaim pada 7 periode berikutnya menggunakan model ARIMA(4,1,0) tanpa konstanta. Pada plot, terlihat bahwa hasil peramalan cukup fluktuatif pada bulan Desember 2023 hingga Juni 2024, dengan tingkat konfidensi 95% berada pada area berwarna biru tua sedangkan tingkat konfidensi 80% berada pada area berwarna biru muda.

#### 4.5. Peramalan dengan *Long Short-Term Memory* (LSTM)

Dalam melakukan peramalan menggunakan machine learning *Recurrent Neural Network* (RNN) seperti *Long Short-Term Memory*, data harus melalui proses *splitting* terlebih dahulu. Setelah melalui proses *splitting*, akan dilakukan pemodelan runtun waktu dengan metode *Long Short-Term Memory* dengan beberapa jumlah *hidden layer*, yakni 1 hingga 4 *hidden layer*.

##### 4.5.1. Data *Splitting*

*Splitting* data merupakan langkah krusial dalam pengembangan model *machine learning* untuk memastikan bahwa model yang dihasilkan bersifat akurat, general, dan dapat diimplementasikan dalam situasi nyata. Pada penelitian ini, data akan dipisah dengan proporsi perbandingan data *train* (data latihan) dan *test* (data uji) sebesar 70:30.

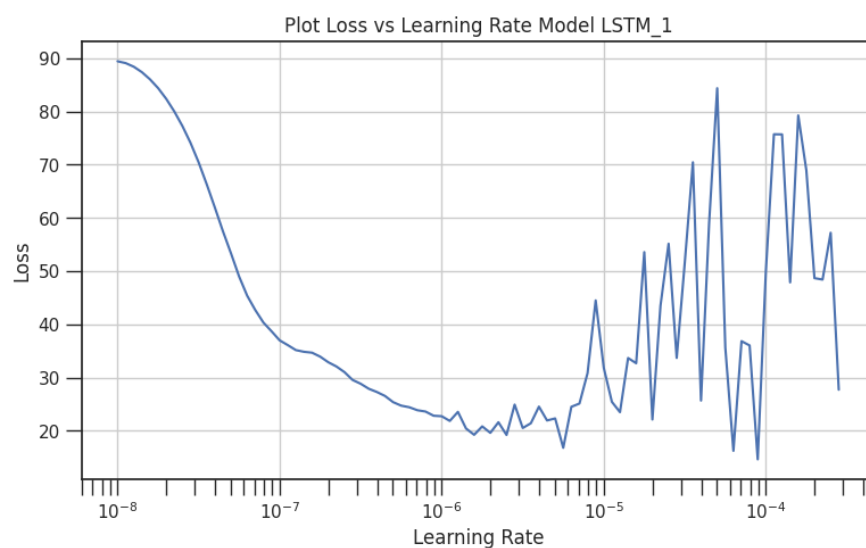


**Gambar 4.13. Plot Hasil *Splitting* Data Besar Klaim**

Data *train* atau latih pada data besar klaim terdiri dari 41 titik data, yang berada pada rentang periode waktu Januari 2019 hingga Mei 2022. Sedangkan data *test* atau uji pada data besar klaim memiliki 18 observasi pada rentang waktu Juni 2022 hingga November 2023. Pada gambar, 4.13, data *train* ditunjukkan dengan garis biru, sementara data *test* direpresentasikan dengan garis jingga. Setelah ini, akan dilakukan pemodelan berdasarkan data *train*. Data *test* akan digunakan untuk menguji akurasi dari model yang sudah terbentuk. Dalam pemodelan pada tahap selanjutnya, optimasi penyesuaian bobot akan dilakukan menggunakan algoritma *Stochastic Gradient Descent* (SGD). Sementara pemilihan *learning rate* paling optimal akan menggunakan 90 *epochs*.

#### 4.5.2. Model LSTM dengan 1 *Hidden Layer*

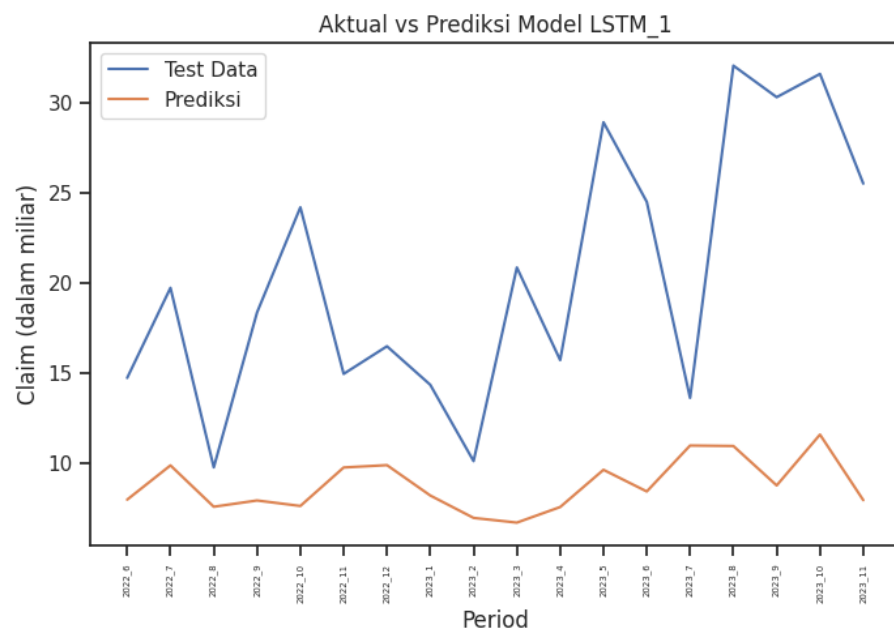
Sebelum melakukan pemodelan *Long Short-Term Memory* menggunakan 1 *hidden layer* terhadap data *train*, akan dilakukan penyesuaian atau tuning terhadap *learning rate* menggunakan algoritma optimasi penyesuaian bobot *Stochastic Gradient Descent*. *Tuning* akan dilakukan menggunakan 90 *epochs* dengan plot *loss* terhadap *learning rate* untuk 1 *hidden layer* sebagai berikut.



**Gambar 4.14. Plot *Loss* terhadap *Learning Rate* Model LSTM\_1**

Pemilihan nilai *learning rate* paling optimal didasarkan pada plot *loss* terhadap *learning rate* yang terbentuk. *Learning rate* yang berada tepat sebelum *loss* mulai menurun tajam merupakan pilihan yang baik, karena hal tersebut menunjukkan bahwa model mulai bekerja lebih efektif. Berdasarkan gambar 4.14, dipilih *learning rate* sebesar  $5,0119 \times 10^{-5}$ .

Selanjutnya, akan dilakukan pelatihan model menggunakan angka *learning rate* yang sudah diperoleh. *Training* model dilakukan dengan optimasi SGD dan *loss function* Huber, menggunakan 500 *epochs* yang disertai dengan *callbacks*. Model yang didapatkan dari proses *training* akan digunakan untuk memprediksi besar klaim. Diperoleh hasil sebagai berikut.



**Gambar 4.15. Plot Data *Test* dan Hasil Peramalan Model LSTM\_1**

```

Test MAPE: 5186.136
Test MAE: 11.526
Test MSE: 176.293
Test RMSE: 13.278

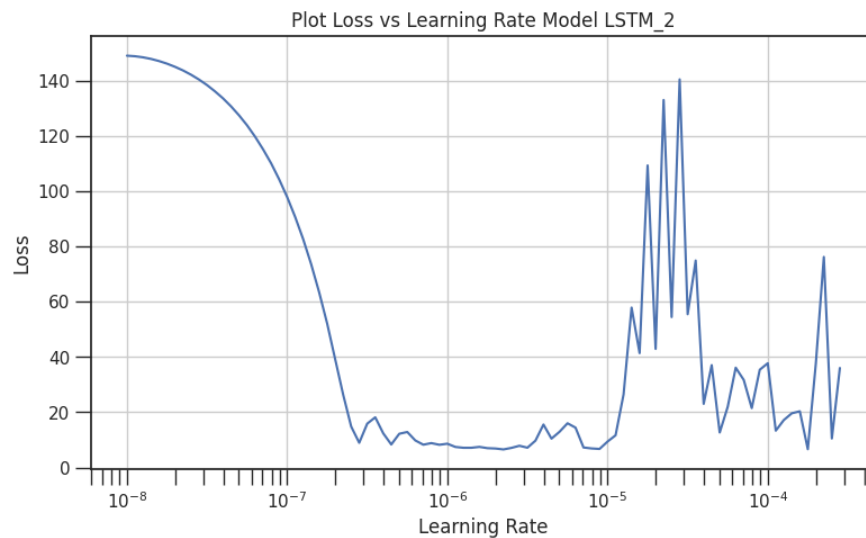
```

**Gambar 4.16. Metrik Evaluasi Hasil Peramalan Model LSTM\_1**

Gambar 4.15 memperlihatkan grafik antara data *test* dan hasil peramalan menggunakan model LSTM dengan 1 *hidden layer*. Akurasi dari hasil *forecast* tersebut ditunjukkan pada gambar 4.16. Diperoleh metrik evaluasi MAE sebesar 11,526, RMSE sebesar 13,278, dan MAPE sebesar 5186,136.

#### 4.5.3. Model LSTM dengan 2 Hidden Layer

Sebelum melakukan pemodelan *Long Short-Term Memory* menggunakan 2 *hidden layer* terhadap data *train*, akan dilakukan penyesuaian atau tuning terhadap *learning rate* menggunakan algoritma optimasi penyesuaian bobot *Stochastic Gradient Descent*. *Tuning* akan dilakukan menggunakan 90 *epochs* dengan plot *loss* terhadap *learning rate* untuk 2 *hidden layer* sebagai berikut.



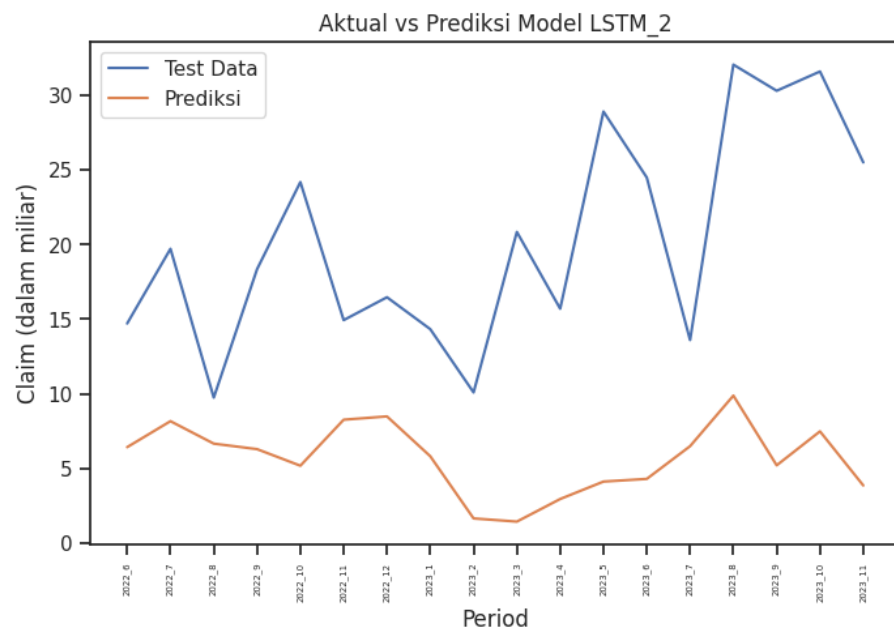
**Gambar 4.17. Plot Loss terhadap Learning Rate Model LSTM\_2**

Pemilihan nilai *learning rate* paling optimal didasarkan pada plot *loss* terhadap *learning rate* yang terbentuk. *Learning rate* yang berada tepat sebelum *loss* mulai menurun tajam merupakan pilihan yang baik, karena hal tersebut menunjukkan bahwa model mulai bekerja lebih



efektif. Berdasarkan gambar 4.17, dipilih *learning rate* sebesar  $2,8184 \times 10^{-5}$ .

Selanjutnya, akan dilakukan pelatihan model menggunakan angka *learning rate* yang sudah diperoleh. *Training* model dilakukan dengan optimasi SGD dan *loss function* Huber, menggunakan 500 *epochs* yang disertai dengan *callbacks*. Model yang didapatkan dari proses *training* akan digunakan untuk memprediksi besar klaim. Diperoleh hasil sebagai berikut.



**Gambar 4.18. Plot Data *Test* dan Hasil Peramalan Model LSTM\_2**

```
Test MAPE: 6856.724
Test MAE: 14.596
Test MSE: 263.787
Test RMSE: 16.242
```

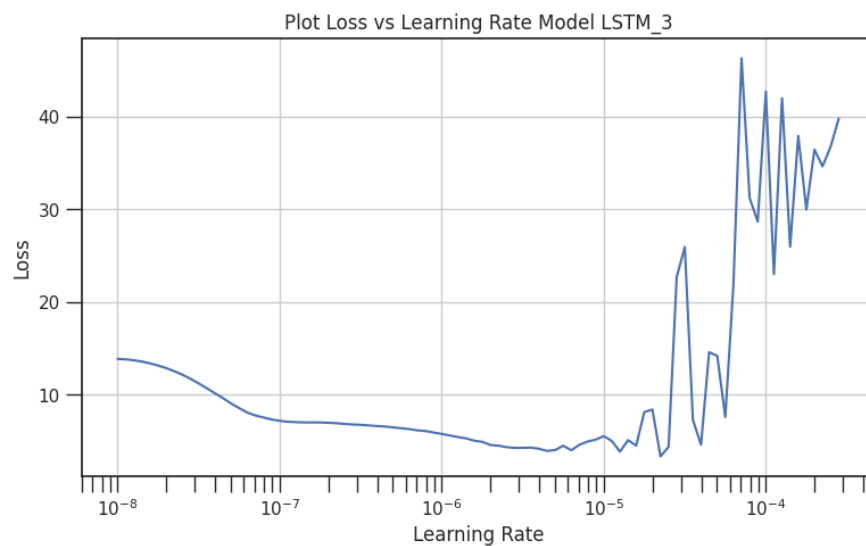
**Gambar 4.19. Metrik Evaluasi Hasil Peramalan Model LSTM\_2**

Gambar 4.18 memperlihatkan grafik antara data *test* dan hasil peramalan menggunakan model LSTM dengan 2 *hidden layer*. Akurasi dari hasil *forecast* tersebut ditunjukkan pada gambar 4.19. Diperoleh

metrik evaluasi MAE sebesar 14,596, RMSE sebesar 16,242, dan MAPE sebesar 6856,724.

#### 4.5.4. Model LSTM dengan 3 *Hidden Layer*

Sebelum melakukan pemodelan *Long Short-Term Memory* menggunakan 3 *hidden layer* terhadap data *train*, akan dilakukan penyesuaian atau tuning terhadap *learning rate* menggunakan algoritma optimasi penyesuaian bobot *Stochastic Gradient Descent*. *Tuning* akan dilakukan menggunakan 90 *epochs* dengan plot *loss* terhadap *learning rate* untuk 3 *hidden layer* sebagai berikut.

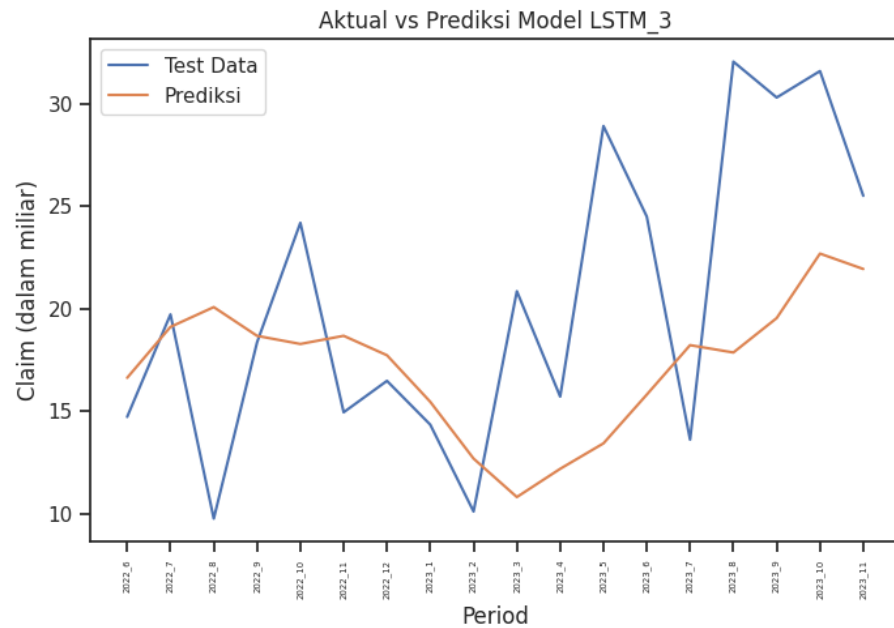


**Gambar 4.20. Plot *Loss* terhadap *Learning Rate* Model LSTM\_3**

Pemilihan nilai *learning rate* paling optimal didasarkan pada plot *loss* terhadap *learning rate* yang terbentuk. *Learning rate* yang berada tepat sebelum *loss* mulai menurun tajam merupakan pilihan yang baik, karena hal tersebut menunjukkan bahwa model mulai bekerja lebih efektif. Berdasarkan gambar 4.20, dipilih *learning rate* sebesar  $3,1623 \times 10^{-5}$ .

Selanjutnya, akan dilakukan pelatihan model menggunakan angka *learning rate* yang sudah diperoleh. *Training* model dilakukan dengan

optimasi SGD dan *loss function* Huber, menggunakan 500 *epochs* yang disertai dengan *callbacks*. Model yang didapatkan dari proses *training* akan digunakan untuk memprediksi besar klaim. Diperoleh hasil sebagai berikut.



**Gambar 4.21. Plot Data *Test* dan Hasil Peramalan Model LSTM\_3**

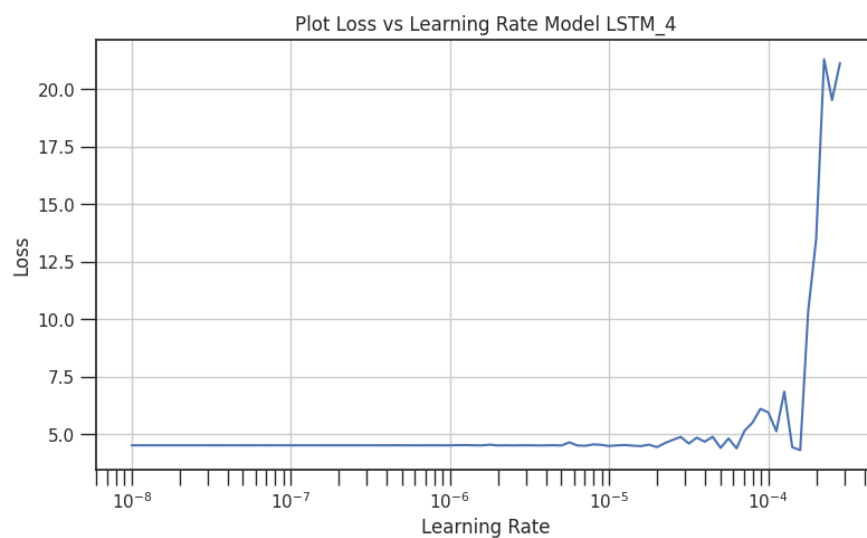
```
Test MAPE: 2941.430
Test MAE: 5.973
Test MSE: 57.106
Test RMSE: 7.557
```

**Gambar 4.22. Metrik Evaluasi Hasil Peramalan Model LSTM\_3**

Gambar 4.21 memperlihatkan grafik antara data *test* dan hasil peramalan menggunakan model LSTM dengan 3 *hidden layer*. Akurasi dari hasil *forecast* tersebut ditunjukkan pada gambar 4.22. Diperoleh metrik evaluasi MAE sebesar 5,973, RMSE sebesar 7,557, dan MAPE sebesar 2941,43.

#### 4.5.5. Model LSTM dengan 4 *Hidden Layer*

Sebelum melakukan pemodelan *Long Short-Term Memory* menggunakan 4 *hidden layer* terhadap data *train*, akan dilakukan penyesuaian atau tuning terhadap *learning rate* menggunakan algoritma optimasi penyesuaian bobot *Stochastic Gradient Descent*. *Tuning* akan dilakukan menggunakan 90 *epochs* dengan plot *loss* terhadap *learning rate* untuk 4 *hidden layer* sebagai berikut.

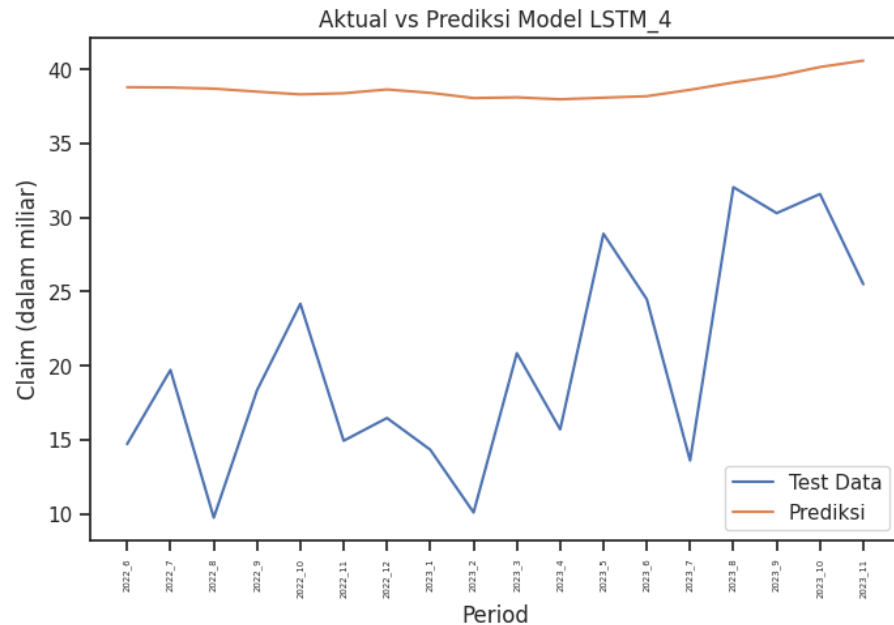


**Gambar 4.23. Plot *Loss* terhadap *Learning Rate* Model LSTM\_4**

Pemilihan nilai *learning rate* paling optimal didasarkan pada plot *loss* terhadap *learning rate* yang terbentuk. *Learning rate* yang berada tepat sebelum *loss* mulai menurun tajam merupakan pilihan yang baik, karena hal tersebut menunjukkan bahwa model mulai bekerja lebih efektif. Berdasarkan gambar 4.23, dipilih *learning rate* sebesar  $1,2589 \times 10^{-4}$ .

Selanjutnya, akan dilakukan pelatihan model menggunakan angka *learning rate* yang sudah diperoleh. *Training* model dilakukan dengan optimasi SGD dan *loss function* Huber, menggunakan 500 *epochs* yang disertai dengan *callbacks*. Model yang didapatkan dari proses *training*

akan digunakan untuk memprediksi besar klaim. Diperoleh hasil sebagai berikut.



**Gambar 4.24. Plot Data *Test* dan Hasil Peramalan Model LSTM\_4**

```
Test MAPE: 11648.032
Test MAE: 18.411
Test MSE: 384.274
Test RMSE: 19.603
```

**Gambar 4.25. Metrik Evaluasi Hasil Peramalan Model LSTM\_4**

Gambar 4.24 memperlihatkan grafik antara data *test* dan hasil peramalan menggunakan model LSTM dengan 4 *hidden layer*. Akurasi dari hasil *forecast* tersebut ditunjukkan pada gambar 4.25. Diperoleh metrik evaluasi MAE sebesar 18,411, RMSE sebesar 19,603, dan MAPE sebesar 11648,032.

#### 4.5.6. Model LSTM Terbaik

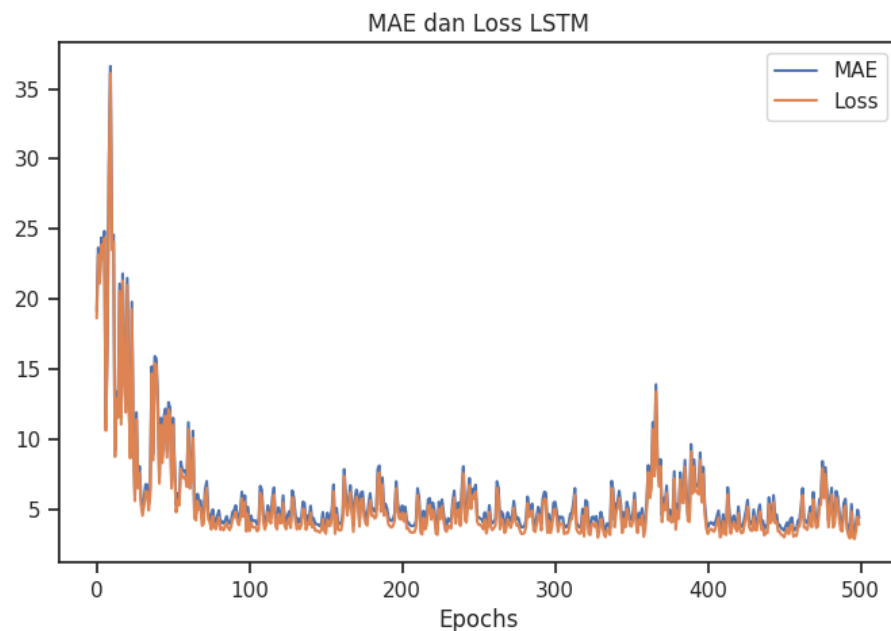
Penentuan model terbaik dilakukan dengan mencermati metrik evaluasi, yaitu MAE, RMSE, dan MAPE. Model yang memiliki nilai MAE, RMSE, dan MAPE terkecil merupakan model LSTM terbaik.

Berikut adalah hasil perhitungan metrik evaluasi tiap model LSTM yang terbentuk.

**Tabel 4.5. Ringkasan Metrik Evaluasi Model LSTM**

<b>Model</b>	<b><i>Hidden layer</i></b>	<b>MAE</b>	<b>RMSE</b>	<b>MAPE</b>
LSTM_1	1	11,526	13,278	5186,136
LSTM_2	2	14,596	16,242	6856,724
LSTM_3	3	5,973	7,557	2941,430
LSTM_4	4	18,411	19,603	11648,032

Berdasarkan tabel 4.5, terlihat bahwa model LSTM\_3 atau model LSTM dengan 3 *hidden layer* memiliki nilai metrik evaluasi terkecil ketika dibandingkan dengan model LSTM lainnya. Maka dapat disimpulkan bahwa model LSTM dengan 3 *hidden layer* merupakan model LSTM terbaik.



**Gambar 4.26. Plot MAE dan *Loss* Model LSTM\_3**

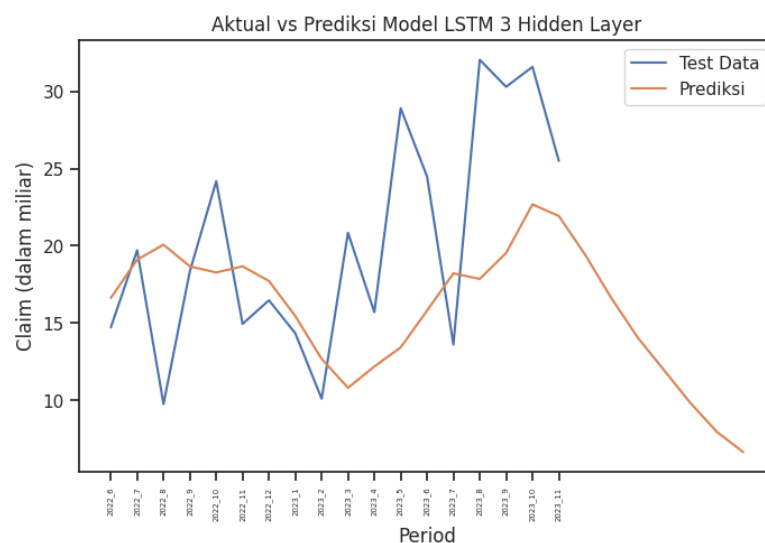
Gambar 4.26 menunjukkan perbandingan antara MAE dan *loss* pada model LSTM\_3, terlihat bahwa MAE dan *loss* bernilai relatif sama,

karena garis biru representasi MAE berdekatan dengan garis jingga representasi *loss*. Nilai MAE dan *loss* yang cenderung menurun seiring dengan bertambahnya jumlah *epochs* memiliki arti bahwa terjadi peningkatan kualitas prediksi model seiring berjalannya proses *training* data. Pada *epochs* atau iterasi awal, model masih memiliki MAE dan *loss* yang besar, namun seiring berjalannya iterasi, model dapat semakin mengoptimalkan parameternya dan memperoleh hasil prediksi yang lebih akurat. Berikut adalah hasil prediksi besar klaim untuk 7 periode berikutnya.

**Tabel 4.6. Hasil Prediksi Besar Klaim dengan Model LSTM\_3**

Periode	Hasil Prediksi Besar Klaim (dalam miliar Rupiah)
Desember 2023	19,424
Januari 2024	16,597
Februari 2024	14,047
Maret 2024	11,930
April 2024	9,806
Mei 2024	7,949

Selanjutnya, akan ditunjukkan hasil prediksi data *test*.



**Gambar 4.27. Plot Hasil Prediksi Model LSTM**

Gambar 4.27 menunjukkan plot perbandingan antara data *test*, hasil prediksi data *test*, dan hasil prediksi untuk 7 periode berikutnya menggunakan metode LSTM dengan 3 *hidden layer*. Terlihat bahwa hasil prediksi besar klaim untuk 7 periode berikutnya cenderung menurun cukup drastis. Penurunan drastis ini ditunjukkan pada plot hasil prediksi model LSTM\_3 berwarna jingga yang menurun dari rentang 20-25 miliar Rupiah menjadi di bawah 10 miliar rupiah.

#### 4.6. Perbandingan Metode

Setelah melakukan pemodelan dan peramalan data besar klaim produk asuransi kesehatan A dan B milik PT X menggunakan metode *Autoregressive Integrated Moving Average* (ARIMA) dan *Long Short-Term Memory* (LSTM), selanjutnya akan dilakukan perbandingan besar metrik evaluasi yang dimiliki dari hasil prediksi kedua model. Metode yang paling baik untuk meramalkan data besar klaim merupakan metode yang memiliki nilai metrik evaluasi *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE) terkecil.

**Tabel 4.7. Ringkasan Metrik Evaluasi Model ARIMA dan LSTM**

Model	MAE	RMSE	MAPE
ARIMA(4,1,0)	4,086	5,654	39,033
LSTM_3	5,973	7,557	2941,430

Berdasarkan tabel 4.7, terlihat bahwa model ARIMA(4,1,0) tanpa konstanta memiliki nilai MAE, RMSE, dan MAPE yang lebih kecil dibandingkan dengan model LSTM dengan 3 *hidden layer*. Maka dapat disimpulkan bahwa metode ARIMA dengan orde AR yaitu 4, orde MA yaitu 0, dan orde diferensi 1 tanpa konstanta merupakan model lebih tepat untuk meramalkan data besar klaim produk asuransi kesehatan A dan B milik PT X jika dibandingkan dengan metode LSTM dengan 3 *hidden layer*.



Berikut adalah perbandingan hasil prediksi besar klaim produk asuransi kesehatan A dan B milik PT X untuk 7 periode berikutnya dengan kedua model yang terbentuk.

**Tabel 4.8. Hasil Prediksi Besar Klaim dengan ARIMA dan LSTM (dalam miliar Rupiah)**

<b>Periode</b>	<b>ARIMA(4,1,0)</b>	<b>LSTM_3</b>
Desember 2023	20,158	19,424
Januari 2024	24,916	16,597
Februari 2024	26,695	14,047
Maret 2024	27,851	11,930
April 2024	26,880	9,806
Mei 2024	24,061	7,949
Juni 2024	24,665	19,424