



PERBANDINGAN MODEL *TIME SERIES FORECASTING* DALAM MEMREDIKSI JUMLAH KEDATANGAN WISATAWAN DAN PENUMPANG AIRPORT

Muhammad Ali Ridla¹, Nur Azise², Miftahur Rahman³

¹ el.riedla@gmail.com, ² nuriza3010@gmail.com, ³ miftahurrahman@unmuhjember.ac.id

^{1,2} Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, ³ Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Abstrak

Kemajuan dunia pariwisata dalam kehidupan zaman sekarang sudah sangat lazim ditemui di setiap negara di dunia. Meningkatkan kualitas pariwisata merupakan hal yang sangat penting bagi setiap negara, mengingat pariwisata merupakan salah satu sumber pemasukan negara. Oleh karena itu, salah satu parameter yang sangat penting akan hal ini adalah mengetahui jumlah pengunjung atau wisatawan setiap waktu, serta dapat memanfaatkan data historis yang ada untuk memprediksi jumlah wisatawan di waktu yang akan datang. Pada penelitian ini akan dilakukan prediksi/forecasting jumlah wisatawan dan penumpang di airport menggunakan metode *Seasonal Auto Regressive Integrated Moving Average* (SARIMA), *Long-short Term Memory* (LSTM), dan Prophet pada dua dataset time series dengan frekuensi bulanan. Dari tiga model forecasting tersebut, diperoleh hasil masing-masing lalu dibandingkan, model SARIMA merupakan model yang paling baik performanya dengan nilai RMSE dan MSE yang paling kecil.

Kata kunci: ARIMA, LSTM, Prediksi, *Time Series*, Univariat

Abstract

The progress of the world of tourism in today's life is very common in every country in the world. Improving the quality of tourism is very important for every country, considering that tourism is a source of state income. Therefore, one of the most important parameters in this regard is knowing the number of visitors or tourists at any time, and being able to utilize existing historical data to predict the number of tourists in the future. This research will predict/forecast the number of tourists and passengers at the airport using the Seasonal Auto Regressive Integrated Moving Average (SARIMA), Long-short Term Memory (LSTM), and Prophet methods on two time series datasets with a monthly frequency. Of the three forecasting models, the results were obtained for each and then compared, the SARIMA model is the model with the best performance with the lowest RMSE and MSE values.

Keywords: ARIMA, LTSM, Predictions, *Time Series*, Univariat

1. Pendahuluan

Kemajuan dunia pariwisata sudah sangat lazim ditemui pada zaman dengan teknologi yang super canggih seperti sekarang ini. Setiap negara pasti akan selalu berusaha meningkatkan kualitas pariwisatanya kepada setiap wisatawan yang berkunjung ke negaranya. Salah satu parameter yang sangat penting akan hal ini adalah mengetahui jumlah pengunjung atau wisatawan setiap waktu [1].

Analisis data jumlah pengunjung ini akan sangat berguna bagi pihak otoritas untuk memperbaiki dan meningkatkan kualitas pariwisatanya. Data-data jumlah wisatawan yang terkumpul dari periode waktu tertentu dapat digunakan untuk memprediksi kejadian-kejadian dimasa depan berdasarkan kejadian-kejadian dimasa lalu, berdasarkan data tersebut [2]. Hal ini akan sangat berguna bagi pihak otoritas dalam melakukan pengambilan keputusan guna meningkatkan kualitas pariwisata di negaranya.

Prediksi tersebut dapat diterapkan pada data *time series* jumlah wisatawan dalam periode tertentu, bisa harian, mingguan, bulanan, hingga tahunan. Ada banyak metode atau model prediksi data *time series* yang dapat digunakan untuk melakukan peramalan (*forecasting*) pada rentang waktu yang akan datang [1].

Seperti penelitian [3][4], yang menggunakan metode gabungan atau yang disebut dengan metode *hybrid* antara metode *Autoregressive Integrated Moving Average* (ARIMA) dan *Artificial Neural Network* (ANN). Ada 3 *dataset* berbeda yang digunakan untuk melakukan eksperimen guna mendemonstrasikan model *hybrid* [3] yaitu: data *Wolf's sunspot*, *Canadian lynx* (kucing lynx) data, dan kurs poundsterling dan dollar Amerika. Hasil dari percobaan ini lalu dikomparasi dengan hasil metode non *hybrid* yaitu ARIMA dan ANN. Metode *hybrid* atau gabungan ARIMA dan ANN berhasil mendapatkan jumlah *Mean Square Error* (MSE) terkecil, hal ini menandakan bahwa model *hybrid* memiliki performa yang lebih baik daripada metode ARIMA dan ANN.

2. Metode

Time series adalah suatu deret data observasi yang berurut berdasarkan waktu [5]. Model *time series* (runtun waktu) telah banyak diimplementasikan dalam berbagai bidang, misalnya teknik, geofisika, ekonomi, pertanian dan kedokteran. *Time series forecasting* adalah proses peramalan atau prediksi kejadian dimasa datang berdasarkan kejadian dimasa lalu menggunakan data historis yang bersifat *time series* (runtun waktu)[6][7]. Berikut akan dijelaskan beberapa metode prediksi (*forecasting*) yang akan digunakan dalam penelitian ini:

2.1. ARIMA dan SARIMA

ARIMA (*Auto Regressive Integrated Moving Average*) atau yang biasa disebut Box-Jenkins merupakan salah satu metode untuk memprediksi data *time series*. Metode ARIMA biasanya digunakan untuk memprediksi kejadian jangka pendek (*short term*)[8]. Adapun metode ARIMA dapat didefinisikan sebagai berikut:

1. Autoregressive (AR)

Metode *Autoregressive* atau disingkat dengan AR adalah metode yang mengatakan bahwa kejadian di periode sekarang dipengaruhi oleh data kejadian di periode sebelumnya [9]. Metode *autoregressive* dengan orde p (AR(p)) atau orde ($p,0,0$) pada model ARIMA dapat dinyatakan sebagai berikut:

$$Y_t = \alpha_0 + \theta_1 Y_{t-1} + \dots + \theta_p Y_{t-p} + e_t \quad (1)$$

Dimana pada parameter Y_t adalah nilai tingkat perbedaan urutan waktu ke- (t) ; α_0 merupakan nilai konstanta. θ_p adalah nilai parameter *auto regressive* ke p . e_t adalah nilai *error* ke (t) .

2. Moving Average (MA)

Moving average (MA) adalah model yang menyatakan bahwa nilai data waktu seluruhnya memiliki hubungan dengan nilai *error* hasil prediksi di masa sekarang dan masa lalu secara berurutan. Model *moving average* dengan orde q disingkat dengan MA(q) atau ARIMA (0,0, q) [7]. Persamaan dasar dari model ini dapat dinyatakan sebagai berikut:

$$Y_t = \theta_0 + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \quad (2)$$

Dimana parameter θ_0 adalah nilai konstanta, θ_q merupakan parameter *moving average* ke q , sedangkan e_{t-p} adalah nilai *error* ke $t - k$.

3. ARMA

Model ARMA atau *Auto Regressive Model Average* adalah model gabungan, dari AR dan MA. Jadi dapat dikatakan bahwa data kejadian di masa sekarang dipengaruhi oleh data periode sebelumnya dan juga kesalahan (*error*) data di masa lalu. Model ARMA dengan orde p dan q disingkat ARMA (p, q) atau ARIMA ($p, 0, q$) [9]. Persamaan dari model ini adalah sebagai berikut:

$$Y_t = \alpha_0 + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} \quad (3)$$

4. ARIMA

Auto Regressive Integrated Moving Average (ARIMA) sama dengan model ARMA namun ada tambahan metode differencing untuk membuat data *time series* menjadi data *stationery*, hal ini dilakukan apabila data *time series*-nya tidak *stationery* [3]. Adapun persamaan ARIMA dengan orde (p, d, q) adalah sebagai berikut:

$$Y_t = (1 + \alpha_1) y_{t-1} + \dots + (1 + \alpha_p) y_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \quad (4)$$

5. SARIMA

Seasonal Auto Regressive Integrated Moving Average adalah model yang sama dengan ARIMA, namun ada tambahan beberapa parameter *seasonality* padanya. Model ini digunakan ketika data *time series* yang diobservasi memiliki pola musiman (*seasonality*) yang berulang-ulang dalam periode waktu tertentu [9]. Adapun persamaannya dapat dibentuk sebagai berikut:

$$ARIMA(p, d, q)(P, D, Q)^s \quad (5)$$

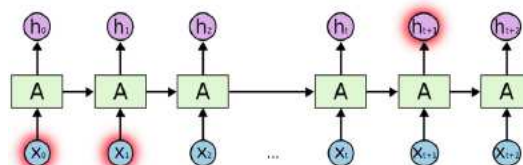
Dimana (p, d, q) dan (P, D, Q) adalah komponen non-musiman dan musiman dari masing-masing model tersebut. Parameter s adalah jumlah periode permusim [10][11]. Penerapan model SARIMA pada penelitian *forecasting* menggunakan persamaan sebagai berikut:

$$\phi(B^s)\varphi(B)(1-B^s)^D(1-B)^d y_t = \theta(B^s)\theta(B)\varepsilon_t \quad (6)$$

Dimana ϕ dan φ adalah parameter *autoregressif* (AR) komponen musiman dan non-musiman, masing-masing; Φ dan Φ parameter *Moving Average* (MA) dari komponen musiman dan non-musiman; B adalah operator mundur, $B(y_t) = y_{t-1}$; $(1-B^s)^D = D^{\text{th}}$ banyaknya perbedaan musim s ; $(1-B)^d = d^{\text{th}}$ perbedaan non-musiman; ε_t adalah variabel acak yang didistribusikan secara independen; P dan p perintah komponen AR; Q dan q perintah komponen MA; D dan d banyaknya perbedaan musim [12].

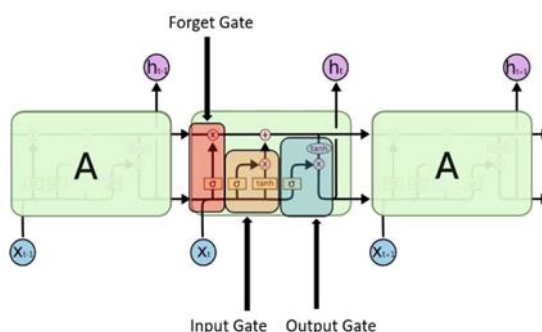
2.2. LSTM

Long Short-term Memory (LSTM) pertama kali dikembangkan oleh Hochreiter dan Schmidhuber (1997) yang memiliki struktur kompleks yang di dalamnya memiliki lapisan tersembunyi [13]. LSTM merupakan teknik turunan dari model *Recurrent Neural Network* (RNN) [14]. RNN merupakan salah satu model yang memiliki performa yang baik untuk menganalisis data-data kejadian yang terjadi dalam waktu yang berurutan (*sequential*), seperti data *time series* [15]. Sedangkan LSTM sendiri memiliki keunggulan tersendiri daripada *simple RNN*, seperti yang ditunjukkan pada gambar 1 di bawah ini:



Gambar 1. Memory di RNN

Pada gambar 1 dijelaskan bahwa kekurangan RNN dapat terlihat pada input X_t , X_{t+1} yang membawa rentang informasi yang sangat besar dengan X_t , X_{t+1} , sehingga ketika output H_{t+1} membutuhkan nilai masukan yang sesuai dengan X_1 , X_0 RNN tidak bias belajar untuk menyesuaikan informasi disebabkan memori yang lama sudah tertimpa dan tergantikan oleh data di memori yang baru dalam rentetan waktu yang berjalan. Namun masalah ini dapat teratasi menggunakan model LSTM yang dapat mengatur memori terhadap setiap input-annya dengan memanfaatkan *memory cells* dan *gate units*, yaitu *Input Gate*, *Forget Gate*, dan *Output Gate*. Seperti yang digambarkan pada gambar 2 ini:



Gambar 2. Memory Cells dan Gate Units LSTM

2.3. Prophet

Model peramalan atau prediksi Prophet, adalah sebuah metode prediksi yang dikembangkan oleh Facebook, yang sudah tersedia dalam Bahasa *R* dan *Python*. Proses prediksinya berdasarkan model tambahan (*additive*) dimana *non-linier* relevan dengan pengaruh *seasonality* seperti tahunan, mingguan, dan harian, serta pengaruh musim liburan. Model *Prophet* ini sangat bagus diterapkan pada data *time series* yang memiliki sifat *seasonality* yang kuat. Model *Prophet* dapat bekerja dengan baik pada data dengan *missing value* dan dapat menangani *outliers* dengan baik. Berikut persamaan model *prophet*:

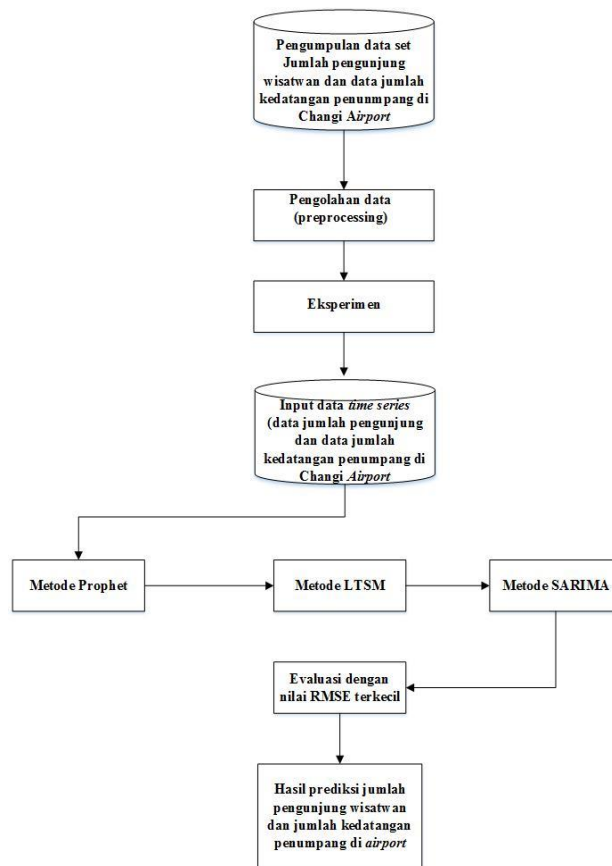
$$y(t) = g(t) + s(t) + \hat{h}(t) + \epsilon_t \quad (7)$$

Dimana parameter-parameter model tersebut $g(t)$, $s(t)$, $h(t)$, ϵ_t : adalah kurva linier bertahap untuk pemodelan perubahan jeda waktu *non-linear* pada data *time series*, perubahan periode waktu, efek pola waktu yang tidak biasa seperti *holiday*, nilai *error* yang dihasilkan dari perubahan yang tidak diakomodasi oleh model secara berurutan [2][16]. Untuk menerapkan model *Prophet forecasting* pada data *time series* yang memiliki *seasonality* dan ingin memprediksi berdasarkan *seasonality* dapat menggunakan deret fourier yang dapat membuat model fleksibel. Pengaruh *seasonality* $s(t)$ dapat dinyatakan dalam persamaan 8 ini [2].

$$s(t) = \sum_{n=1}^N a_n \cos\left(\frac{2\pi nt}{p}\right) + b_n \sin\left(\frac{2\pi nt}{p}\right) \quad (8)$$

2.4. Tahapan Penelitian

Pada penelitian ini dilakukan beberapa tahapan, yaitu dengan melakukan eksperimen pada data jumlah pengunjung wisatawan dan data jumlah kedatangan penumpang di Changi Airport Singapura. Adapun tahapan penyelesaian ditunjukkan pada gambar 3 di bawah ini:



Gambar 3. Alur Tahapan Penyelesaian

2.5. Pengumpulan Data

Metode pengumpulan data pada penelitian ini menggunakan data *dataset* global yang diperoleh dari www.kaggle.com, ada dua *dataset time series*, yaitu:

1. Data jumlah pengunjung negara Taiwan dengan frekuensi waktu bulanan dari tahun 2011 sampai tahun 2018, ada 96 data record. Pada *dataset* ini, peneliti hanya memilih jumlah pengunjung dari negara Indonesia saja, untuk lebih menyederhanakan proses analisis.
2. Data jumlah kedatangan penumpang di Changi Airport Singapura, dengan frekuensi waktu bulanan dari tahun 2009 sampai tahun 2019, ada 132 data record.

Kedua *dataset* tersebut merupakan data *time series univariate* yang memiliki frekuensi waktu bulanan.

2.6. Pengolahan data

Sebelum melakukan eksperimen dan pengujian data, tahap yang harus dilakukan adalah melakukan pengolahan data pada tahap *preprocessing*. Pengolahan data dilakukan pada masing-masing metode, yaitu Pada metode SARIMA tahap *preprocessing* pada metode SARIMA menggunakan metode *differencing*. Sedangkan pada model model LSTM, dataset terlebih dahulu dinormalisasi guna mengurangi tingkat nilai *error*. Teknik normalisasi *min-max scaling* akan digunakan pada kedua dataset tersebut. *Min-max scaling* adalah mengubah data real menjadi nilai range interval dengan persamaan berikut [17]:

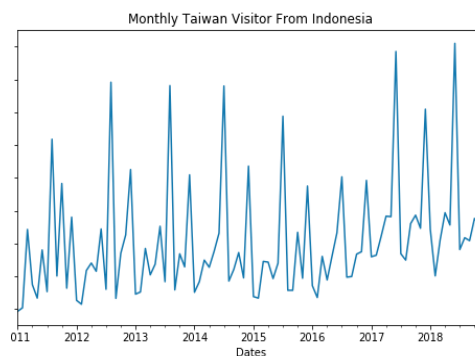
$$\hat{x}_{ik} = \frac{x_{ik} - \min(x_k)}{\max(x_k) - \min(x_k)} \quad (9)$$

2.7. Eksperimen

Tahap ini merupakan tahap pengujian pada masing-masing data, yaitu data jumlah pengunjung wisatawan dan data kedatangan penumpang di Changi Airport Singapura yang dilakukan dengan menggunakan metode *Prophet*, *LSTM*, dan *SARIMA* untuk mendapatkan nilai *MSE* dan *RMSE* terkecil.

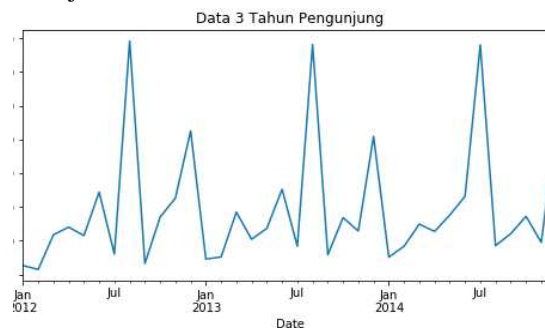
3. Hasil dan Pembahasan

Sebelum melakukan *forecasting*, analisis data diperlukan guna mendapatkan *insight* dari data *time series* tersebut. Berikut akan ditampilkan *line plot diagram* dari kedua *dataset* tersebut:



Gambar 4. Data Pengunjung Taiwan dari Indonesia

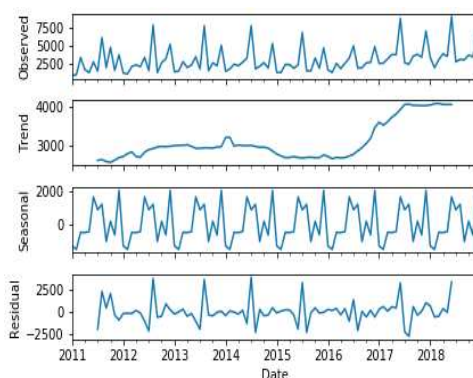
Dari *line plot diagram* tersebut, dapat dilihat jumlah pengunjung cukup fluktuatif setiap tahunnya, dan memiliki hampir pola *seasonality* di setiap tahunnya. Untuk lebih jelasnya, berikut *line plot diagram* dengan *sample* periode 3 tahun saja.



Gambar 5. Data Pengunjung Taiwan dari Indonesia tahun 2012-2014

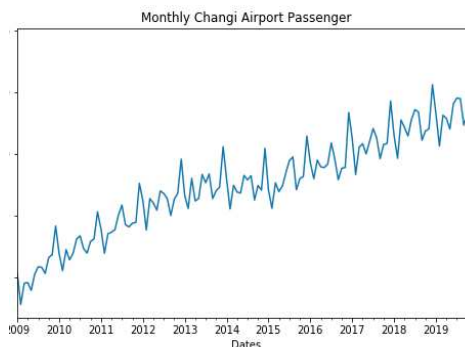
Terlihat cukup jelas fluktuasi jumlah pengunjung mengalami pola *seasonality* yang sama setiap tahunnya. Jumlah pengunjung selalu meningkat di bulan Agustus dan Desember hal ini diperkirakan karena liburan sekolah akhir semester, dan juga liburan akhir tahun. Namun jumlah pengunjung akan kembali menurun di bulan Februari dan Januari.

Setiap data *time series* harus setidaknya memiliki tiga unsur ini; *Trend*, *Seasonality*, dan *Residual*. Berikut *line diagram* dari ketiga unsur tersebut pada *dataset* jumlah pengunjung Taiwan dari Indonesia:

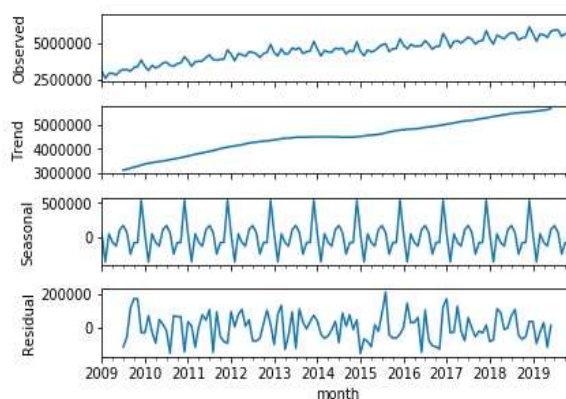


Gambar 6. Tren, Seasonal, dan Residual Pengunjung Taiwan dari Indonesia

Data tersebut memiliki tren yang cukup stabil di tahun 2011-2016, dan memiliki tren yang meningkat (*upward*) di tahun 2016-2018. *Seasonality* data tersebut juga cukup terlihat signifikan.



Gambar 7. Data Jumlah Kedatangan Penumpang di Changi Airport Singapura



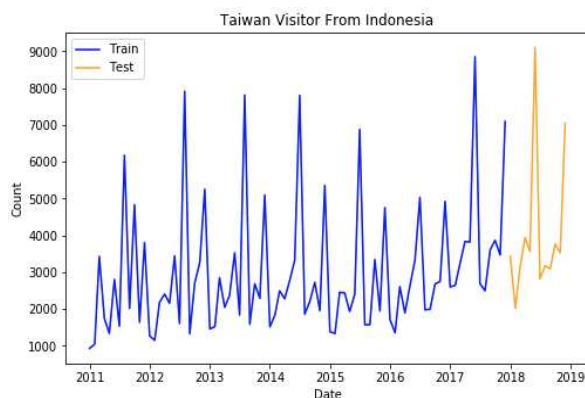
Gambar 8. Tren, Seasonal, dan Residual Jumlah Kedatangan Penumpang di Changi Airport Singapura

Tidak jauh berbeda dengan data pengunjung di Taiwan, pada dataset jumlah kedatangan penumpang di Changi *airport* juga cukup fluktuatif. Terdapat pola *seasonality* yang terjadi hampir setiap tahunnya, jumlah kedatangan penumpang meningkat di bulan Juli dan Desember, dan kembali turun di bulan Februari. Tren yang dimiliki cukup kuat, disini terdapat tren meingkat (*upward*).

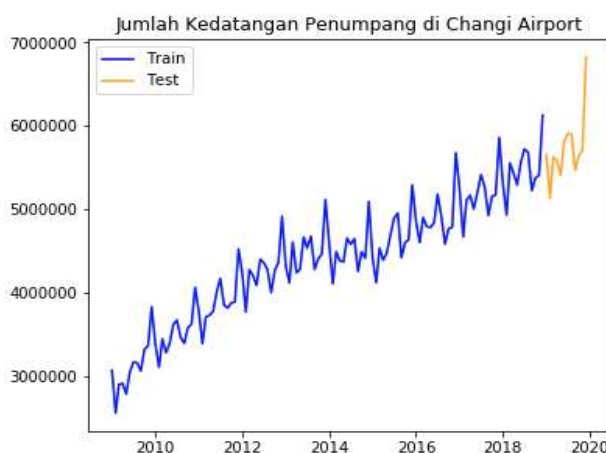
Pada tahap implementasi dan hasil model *time series forecasting* ini akan dilakukan percobaan *forecasting* pada kedua dataset tersebut. Model yang akan digunakan adalah SARIMA, LSTM (RNN) dan

Prophet Forcasting Model. Dalam implementasi ini semua dataset akan dibagi menjadi dua, yaitu data latih, dan data uji guna memvalidasi hasil implementasi model pada data latih.

Untuk dataset jumlah pengunjung Taiwan dari Indonesia, data latihnya berjumlah 84 (dalam rentang waktu 2011-2017), dan data ujinya berjumlah 12 (1 tahun terakhir, 2018). Sedangkan untuk dataset jumlah kedatangan penumpang di Changi *Airport*, data latih berjumlah 120 (dalam rentang waktu 2009-2018), dan data uji berjumlah 12 (1 tahun terakhir 2019). Berikut *line plot* data latih dan data uji dari kedua dataset tersebut:



Gambar 9. Data latih dan data uji dataset jumlah pengunjung Taiwan dari Indonesia



Gambar 10. Data latih dan data uji dataset jumlah Kedatangan Penumpang di Changi *Airport* Singapura

3.1. Implementasi Hasil SARIMA

Pada Model Seasonal Auto Integrated Moving Average ini, *dataset* tersebut haruslah bersifat *stationary*. Jika data tersebut tidak *stationary* dapat dilakukan sebuah preprocessing yaitu dengan metode *differencing*. Oleh karena itu, untuk mengetahui dataset tersebut *stationary* atau tidak, dapat dilakukan *stationary test* yang salah satunya dengan teknik *Dickey Fuller Statistical Test*. Berikut hasil dari *Dickey Fuller Statistical Test* tersebut:

Tabel 1 ADF Test Data Pengunjung Taiwan dari Indonesia

Dataset Pengunjung Taiwan dari Indonesia	
Dickey Fuller Statistical Test	0
P-value	1
Used Lags	12
Number of comments used	83
Critical Value (1%)	-3
Critical Value (5%)	-3
Critical Value (10%)	-3
Dickey Fuller Statistical Test	0

Tabel 2 ADF Test Jumlah Kedatangan Penumpang di Changi Airport Singapura

Dataset Jumlah Kedatangan Penumpang di Changi Airport Singapura	
Dickey Fuller Statistical Test	-0
P-value	1
Used Lags	12
Number of comments used	119
Critical Value (1%)	-3
Critical Value (5%)	-3
Critical Value (10%)	-3

Berdasarkan hasil ADF tersebut, semua hasil statistic ADF lebih besar dari critical values, jadi dapat disimpulkan bahwa semua dataset bersifat non-stationary. Hasil test statistik masih juga harus dicocokkan atau dikonfirmasi dengan teknik visual pada diagram line-nya dan harus dilakukan percobaan pada nilai masukan Integrated (I/ Differencing) saat mengimplementasikan model SARIMA.

Selanjutnya akan dilakukan percobaan menggunakan metode SARIMA, dengan urutan masukan ARIMA $(p,d,q)(P,D,Q)s$ seperti yang dijelaskan pada sub bab 11.1 sebelumnya. Nilai masukan tersebut akan dicoba dengan nilai *differencing* yang berbeda lalu akan dikomparasi hasil RMSE dan MSE nya pada data latih.

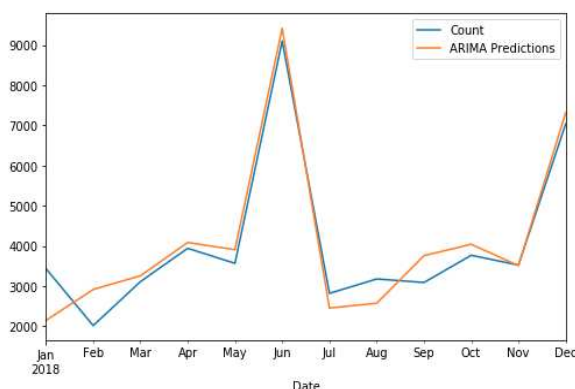
1. Model SARIMA pada dataset pengunjung Taiwan dari Indonesia

Hasil eksperimen dengan model SARIMA pada dataset pengunjung di Taiwan dari Indonesia ditunjukkan pada tabel 3 di bawah ini.

Tabel 3 Hasil SARIMA Forecasting

$(p,d,q)(P,D,Q)s$	MSE	RMSE
(1,0,0)(1,1,1)12	319425,29	565,18
(1,1,0)(1,1,1)12	2926073,47	1710,58

Dari hasil percobaan pada data latih lalu dilakukan validasi pada data uji, didapatkan nilai masukan ARIMA (1,0,0)(1,1,1)12 memiliki nilai MSE dan RMSE yang paling kecil. Berikut line plot hasil forecasting menggunakan model SARIMA pada data latih:



Gambar 11. Perbandingan data uji dengan hasil prediksi SARIMA (1,0,0)(1,1,1)12

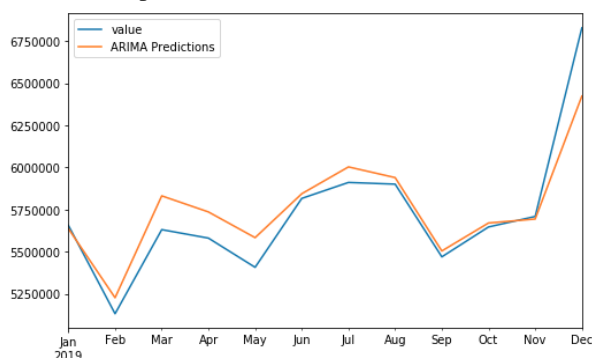
2. Model SARIMA pada dataset jumlah kedatangan penumpang di Changi Airport

Hasil eksperimen dengan model SARIMA pada dataset pengunjung di Taiwan dari Indonesia ditunjukkan pada tabel 3 di bawah ini.

Tabel 4 Hasil SARIMA Forecasting

$(p,d,q)(P,D,Q)s$	MSE	RMSE
$(0,0,1)(2,1,2)12$	29.657.018.261	172212,1316
$(0,1,1)(2,1,2)12$	23.452.430.508	153141,864

Dari hasil percobaan pada data latih lalu dilakukan validasi pada data uji, didapati bahwa nilai masukan ARIMA (0,1,1)(2,1,2)12 memiliki nilai MSE dan RMSE yang paling kecil. Berikut line plot hasil forecasting menggunakan model SARIMA pada data latih:



Gambar 12 Perbandingan data uji dengan hasil prediksi SARIMA (0,1,1)(2,1,2)12

3.2. Implementasi LSTM RNN

1. Preprocessing

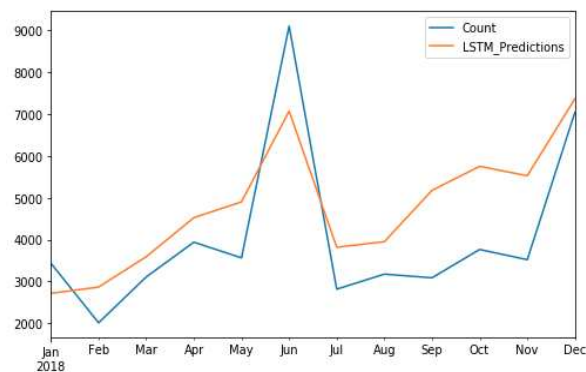
Sebelum melakukan percobaan model LSTM, dataset terlebih dahulu dinormalisasi guna mengurangi tingkat nilai error. Teknik normalisasi *min-max scaling* akan digunakan pada kedua dataset tersebut. *Min-max scaling* adalah mengubah data real menjadi nilai *range interval* [0,1].

2. Long-short Term Memory (LSTM) RNN

Setelah melakukan *preprocessing* selanjutnya akan dilakukan percobaan *forecasting* menggunakan model LSTM. Pada LSTM disini, menggunakan 1 hidden layer dengan 200 dimensi, dan menggunakan activation function ReLu, serta Epoch= 100. Berikut hasil MSE dan RMSE dari LSMTM pada kedua dataset masing-masing.

Tabel 5 Hasil LSTM *Forecasting* Jumlah Pengunjung di Taiwan dari Indonesia

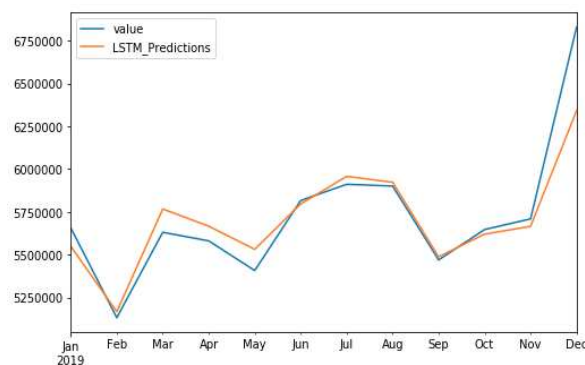
LSTM: Dataset Jumlah Pengunjung Taiwan dari Indonesia	
MSE	RMSE
1813932,153	1346,822985



Gambar 13 Perbandingan data uji dengan hasil prediksi LSTM Dataset Pengunjung Taiwan

Tabel 6 Hasil LSTM *Forecasting* Jumlah Kedatangan Penumpang di Changi Airport

LSTM: Dataset Jumlah Kedatangan Penumpang di Changi <i>Airport</i>	
MSE	RMSE
24.807.219.624	157503,0781



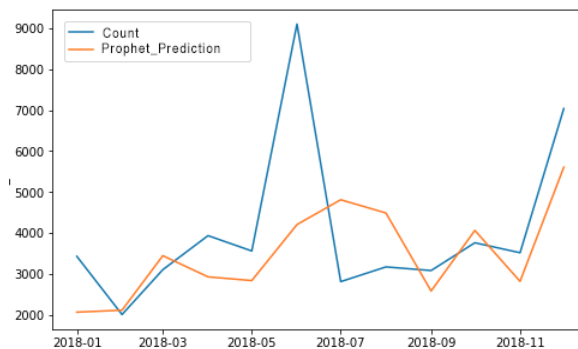
Gambar 14 Perbandingan data uji dengan hasil prediksi LSTM Dataset Kedatangan Changi *Airport*

3.3. Implementasi *Prophet Forecasting Model*

Berikut hasil MSE & RMSE dari percobaan *forecasting* menggunakan model *Prophet*, dengan masukan pada persamaan yang sudah dijelaskan pada sub bab 2.3.

Tabel 7 Hasil Prophet *Forecasting*

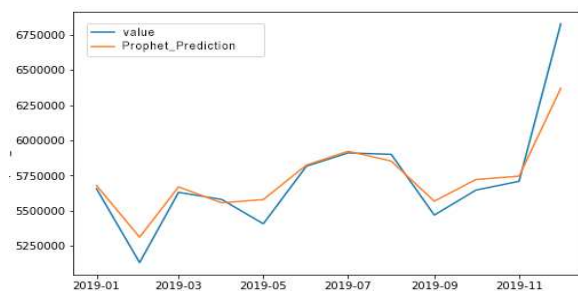
<i>Prophet</i> : Dataset Jumlah Pengunjung Taiwan dari Indonesia	
MSE	RMSE
3003489,275	1733,057782



Gambar 15 Perbandingan data uji dengan hasil prediksi *Prophet* Dataset Pengunjung Taiwan

Tabel 8 Hasil Prophet Forecasting

<i>Prophet: Dataset Jumlah Kedatangan Penumpang di Changi Airport</i>	
MSE	RMSE
24.558.451.552	156711,3638



Gambar 16 Perbandingan data uji dengan hasil prediksi *Prophet* Dataset Kedatangan Changi Airport

Berikut adalah hasil komparasi dari hasil semua percobaan pada tiga model *forecasting* yaitu SARIMA, LSTM, dan *Prophet*:

Tabel 9 Komparasi Tiga Model Forecasting

Dataset Jumlah Pengunjung Taiwan dari Indonesia		
Model	RMSE	MSE
SARIMA	565	319.425
LSTM	1.347	1.813.932
<i>Prophet</i>	1.733	3.003.489

Tabel 10 Perbandingan Data Real dan Hasil Forecasting Data Uji Pengunjung Taiwan dari Indonesia

Month	Real	ARIMA	LSTM	<i>Prophet</i>
01/18	3434	2142	2715	2073
02/18	2014	2914	2867	2120
03/18	3109	3255	3592	3450
04/18	3939	4084	4525	2933
05/18	3565	3904	4902	2843
06/18	9101	9415	7071	4209
07/18	2817	2452	3819	4815
08/18	3176	2571	3951	4491
09/18	3088	3756	5177	2588
10/18	3765	4039	5752	4067
11/18	3522	3508	5526	2824
12/18	7039	7316	7363	5608

Tabel 11 Komparasi Tiga Model Forecasting
Dataset Jumlah Kedatangan Penumpang di Changi Airport

Model	RMSE	MSE
SARIMA	153.142	23.452.430.508
LSTM	157.503	24.807.219.624
Prophet	156.711	24.558.451.552

Tabel 12 Perbandingan Data Real dan Hasil Forecasting pada Data Uji Jumlah Kedatangan Changi Airport Singapura

Month	Real	ARIMA	LSTM	Prophet
01/19	5.656.076	5.636.642	5.548.514	5.679.077
02/19	5.131.807	5.226.755	5.166.953	5.312.444
03/19	5.630.780	5.831.258	5.766.676	5.669.597
04/19	5.580.503	5.735.820	5.666.081	5.556.350
05/19	5.407.308	5.583.248	5.531.086	5.579.520
06/19	5.816.089	5.843.826	5.796.431	5.824.505
07/19	5.910.782	6.002.966	5.956.819	5.922.897
08/19	5.900.629	5.939.459	5.922.748	5.852.532
09/19	5.469.342	5.504.248	5.487.570	5.567.841
10/19	5.646.643	5.670.161	5.619.881	5.721.932
11/19	5.708.993	5.693.691	5.665.265	5.745.476
12/19	6.827.843	6.422.891	6.340.182	6.368.971

Berdasarkan hasil pengujian pada Data jumlah pengunjung negara Taiwan dan data jumlah kedatangan penumpang di Changi Airport Singapura, model SARIMA menunjukkan hasil yang lebih baik, hal tersebut juga ditunjukkan pada hasil penelitian [18][19].

4. Kesimpulan dan Saran

Prediksi atau *forecasting* pada kedua dataset time series; jumlah pengunjung Taiwan dari Indonesia, dan jumlah kedatangan penumpang di Changi Airport menggunakan tiga model *forecasting* yaitu SARIMA, LSTM, dan Prophet, dapat diketahui dari hasil komparasi nilai RMSE dan MSE yang diujikan pada data uji, model SARIMA memiliki performa yang paling baik di antara tiga model tersebut, dengan nilai RMSE dan MSE yang paling kecil.

Daftar Pustaka

- [1] J. L. Chen, G. Li, D. C. Wu, and S. Shen, "Forecasting Seasonal Tourism Demand Using a Multiseries Structural Time Series Method," *J. Travel Res.*, vol. 58, no. 1, pp. 92–103, 2019, doi: 10.1177/0047287517737191.
- [2] K. K. R. Samal, K. S. Babu, S. K. Das, and A. Acharaya, "Time series based air pollution forecasting using SARIMA and prophet model," in *ACM International Conference Proceeding Series*, 2019, pp. 80–85. doi: 10.1145/3355402.3355417.
- [3] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," in *Neurocomputing*, 2003, vol. 50, pp. 159–175. doi: 10.1016/S0925-2312(01)00702-0.
- [4] D. Ömer Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction," *Eng. Appl. Artif. Intell.*, vol. 23, no. 4, pp. 586–594, 2010, doi: 10.1016/j.engappai.2009.09.015.
- [5] K. Bandara, C. Bergmeir, and S. Smyl, "Forecasting Across Time Series Databases using Recurrent Neural Networks on Groups of Similar Series: A Clustering Approach," *Expert Syst. Appl.*, vol. 140, Oct. 2019, doi: 10.1016/j.eswa.2019.112896.
- [6] A. Juliana, Hamidatun, and R. Muslima, *Modern Forecasting Teori dan Aplikasi (GARCH, Artificial Neural Network, Neuro-Garch)*. Yogyakarta: Deepublish, 2019.
- [7] A. Cyril, R. H. Mulangi, and V. George, "Modelling and Forecasting Bus Passenger Demand using <http://e-jurnal.stmikbinsa.ac.id/index.php/simkom>

-
- Time Series Method,” *2018 7th Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2018*, vol. 58, pp. 460–466, 2018, doi: 10.1109/ICRITO.2018.8748443.
- [8] M. Ekananda, *Analisis Data Time Series*, 1st ed. Jawa Barat: Mitra Wacana Media, 2014.
- [9] Z. Li, J. Bi, and Z. Li, “Passenger Flow Forecasting Research for Airport Terminal Based on SARIMA Time Series Model,” 2017. doi: 10.1088/1755-1315/100/1/012146.
- [10] S. Noureen, S. Atique, V. Roy, and S. Bayne, “Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small scale agricultural load,” *Midwest Symp. Circuits Syst.*, vol. 2019-Augus, pp. 521–524, 2019, doi: 10.1109/MWSCAS.2019.8885349.
- [11] P. Chen, A. Niu, D. Liu, W. Jiang, and B. Ma, “Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing,” in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 394, no. 5. doi: 10.1088/1757-899X/394/5/052024.
- [12] K. B. Tadesse and M. O. Dinka, “Application of SARIMA model to forecasting monthly flows in Waterval River, South Africa,” *J. Water L. Dev.*, vol. 35, no. 1, pp. 229–236, 2017, doi: 10.1515/jwld-2017-0088.
- [13] A. Soy Temür, M. Akgün, and G. Temür, “Predicting housing sales in turkey using arima, lstm and hybrid models,” *J. Bus. Econ. Manag.*, vol. 20, no. 5, pp. 920–938, 2019, doi: 10.3846/jbem.2019.10190.
- [14] Z. Zhao, W. Chen, X. Wu, P. C. V. Chen, and J. Liu, “LSTM network: A deep learning approach for short-term traffic forecast,” *IET Image Process.*, vol. 11, no. 1, pp. 68–75, 2017, doi: 10.1049/iet-its.2016.0208.
- [15] K. Chen, Y. Zhou, and F. Dai, “A LSTM-based method for stock returns prediction: A case study of China stock market,” in *2015 IEEE International Conference on Big Data (Big Data)*, Oct. 2015, pp. 2823–2824. doi: 10.1109/BigData.2015.7364089.
- [16] S. J. Taylor and B. Letham, “Forecasting at Scale,” *Am. Stat.*, vol. 72, no. 1, pp. 37–45, 2018, doi: 10.1080/00031305.2017.1380080.
- [17] E. Prasetyo, *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi, 2014.
- [18] L. Nashold and R. Krishnan, “Using LSTM and SARIMA Models to Forecast Cluster CPU Usage,” 2020.
- [19] S. Alkharif, K. Lee, and H. Kim, “Time-Series Analysis for Price Prediction of Opportunistic Cloud Computing Resources,” *Lect. Notes Electr. Eng.*, vol. 461, pp. 221–229, 2018, doi: 10.1007/978-981-10-6520-0_23.