

Penggunaan Variabel Event dan Libur Sekolah Dalam Memprediksi Wisatawan Dengan Metode LSTM

Candra Rusmana¹, Kusrini², Kusnawi³

^{1,2,3}Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

¹candrarusmana@students.amikom.ac.id, ²kusrini@amikom.ac.id, ³khusnawi@amikom.ac.id

Abstract

Events held in regions and School holiday are annuals agenda for families to go on vacations. The fluctuation of tourists has implications for the government, communities, and industry stakeholders. The aim of this research is to examine the effects of annual events and school holidays as variables. The dataset used is from website of data.ntbprov.go.id. The dataset of annual events is from disbudpar.ntbprov.go.id, while the dataset for school holiday from kalenderpendidikan.com. The data collected from 2017 to 2022. For prediction, the Long Short-Term Memory algorithm is employed using the variables of historical, events, and holidays. The utilization of historical, event, and holiday variables yields a MAPE performance of 20.8% using an 90/10 split for training and testing data. The results with historical and holiday variables produce a MAPE of 38.6%, using historical and event variables results in a MAPE of 23.81%. This research introduces a novel approach to predicting the number of tourists by utilizing annual event variables and holiday through the implementation of the LSTM algorithm. In conclusion, this study highlights the importance of considering annual events and holidays as variables for predicting tourist. This research contributes by introducing a new approach that combines event variables and holiday for tourism prediction using LSTM algorithm.

Keywords: event, holidays, LSTM, prediction, tourists

Abstrak

Event yang diadakan diberbagai daerah menjadi daya tarik tersendiri bagi wisatawan untuk datang ketempat tersebut. Liburan musiman sekolah juga menjadi agenda tahunan keluarga untuk pergi ke tempat wisata. Naik turunnya jumlah wisatawan yang datang ke Provinsi NTB memberikan dampak kepada pemerintah daerah, masyarakat sekitar tempat wisata dan pelaku usaha bidang pariwisata. Tujuan penelitian ini untuk melakukan pengujian terhadap variabel event tahunan dan libur sekolah. *Dataset* yang digunakan didapatkan dari website publik Provinsi NTB yaitu data.ntbprov.go.id *dataset* tersebut berupa histori jumlah kunjungan wisatawan setiap bulan, dari website disbudpar.ntbprov.go.id didapatkan *dataset* event tahunan dan dari website kalenderpendidikan.com didapat *dataset* kalender akademik untuk liburan sekolah, *dataset* yang diambil dari setiap sumber diambil mulai dari tahun 2017 sampai tahun 2022. Dari semua *dataset* yang didapat bisa dimanfaatkan dalam menggali informasi untuk melakukan prediksi. Dalam melakukan prediksi digunakan Algoritma LSTM dengan menggunakan variabel histori wisatawan, event dan libur sekolah. Penggunaan variabel histori, event dan liburan menghasilkan kinerja MAPE sebesar 20.8% dengan penggunaan *data training* dan *data testing* 90/10. Hasil kinerja dengan variabel histori dan liburan saja menghasilkan kinerja MAPE sebesar 38,6%. sedangkan hasil dengan variabel histori dan event saja menghasilkan kinerja MAPE sebesar 23,81%. Ini menunjukkan bahwa variabel event dan kalender liburan bisa dengan baik digunakan dalam melakukan prediksi terhadap kedatangan wisatawan di waktu berikutnya. Penelitian ini memperkenalkan pendekatan baru dalam memprediksi jumlah wisatawan dengan menggunakan variabel event tahunan dan kalender libur sekolah dengan menggunakan algoritma LSTM sebagai alat prediksi yang lebih canggih, yang sebelumnya belum banyak dieksplorasi dalam konteks prediksi pariwisata di Provinsi NTB.

Kata kunci: *event*, liburan, LSTM, prediksi, wisatawan

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Pariwisata merupakan salah satu sektor penting dalam perekonomian suatu daerah [1]. Untuk meningkatkan daya tarik pariwisata di suatu daerah dan dalam mengoptimalkan pengelolaan destinasi wisata, penting untuk memahami dan melakukan prediksi jumlah wisatawan yang akan berkunjung pada suatu periode waktu tertentu [2][3]. Prediksi yang akurat tentang jumlah wisatawan dapat membantu dalam perencanaan, pengelolaan sumber daya, dan pengambilan keputusan strategis yang berkaitan dengan industri pariwisata [4].

Saat ini tersedia berbagai jenis data yang dapat digunakan dalam memprediksi jumlah wisatawan. Salah satu sumber data yang akan digunakan dalam melakukan prediksi pada penelitian ini adalah *dataset event* [4], yang mencakup informasi tentang acara tahunan dengan skala nasional dan internasional, festival daerah tahunan, atau kegiatan daerah lainnya yang dapat menarik minat wisatawan. Selain itu, kalender sekolah juga dapat memberikan wawasan yang berharga, karena periode liburan sekolah sering kali menjadi waktu yang sibuk bagi destinasi pariwisata sehingga data libur sekolah bisa dijadikan pengetahuan dalam melakukan prediksi jumlah wisatawan di Prov. NTB.

Dalam penelitian ini, peneliti akan menjelajahi penggunaan *dataset event* dan kalender sekolah dalam memprediksi jumlah wisatawan. Peneliti akan menganalisis hubungan antara jenis acara atau kegiatan tertentu serta liburan sekolah dengan peningkatan kunjungan wisatawan.

Tujuan dari penelitian ini adalah untuk menggali pengetahuan tentang penggunaan *dataset event* dan kalender libur sekolah dalam melakukan model prediksi jumlah wisatawan yang di Provinsi Nusa Tenggara Barat. Dari hasil penggalian pengetahuan tersebut kedepannya dapat digunakan bagi pihak-pihak terkait dalam industri pariwisata. Dengan menggunakan *dataset event* dan kalender sekolah, kami berharap dapat memberikan panduan yang lebih baik dalam perencanaan dan pengelolaan destinasi wisata, sehingga meningkatkan pengalaman wisatawan dan memaksimalkan potensi ekonomi yang dihasilkan oleh sektor pariwisata [5].

Dalam penelitian ini metode yang akan digunakan dalam menguji *dataset event* dan kalender libur sekolah adalah metode *Long Short Time Series* (LSTM). Beberapa penelitian dalam melakukan prediksi wisatawan menggunakan metode *Long Short Time Series* (LSTM) sudah banyak dilakukan, seperti penelitian yang dilakukan oleh (Jian-Wu Bi, Yang Liu, dan Hui Li, 2020) dengan penelitian yang berjudul *Daily Tourism Volume Forecasting For Tourist Attractions* menghasilkan pengujian data historis, mesin pencari dan cuaca menghasilkan kekuatan peramalan yang sangat bagus dibandingkan dengan hanya menggunakan satu variabel saja, hal ini menunjukkan bahwa *dataset* mesin pencari dan *dataset* cuaca sangat baik dan penting untuk digunakan dalam peramalan wisatawan [6].

Selanjutnya penelitian (Harun Mukhtar, Yoze Rizki, Febby Apri Wenando, Muhammad Abdul Al Aziz, 2022) tentang Prediksi Kunjungan Wisatawan dengan *Reduksi Noise* pada *Google Trends* menggunakan *Hilbert-Huang Transform* dan *Long Short-Term Memory* menghasilkan RMSE LSTM 129249 dan RMSE HHT + LSTM 653058 dengan jumlah *dataset* 156 baris dan jumlah neuron 100 unit [7].

Selanjutnya penelitian yang dilakukan oleh (YiFei Li dan Han Cao, 2018) tentang *Prediction for Tourism Flow Based on LSTM Neural Network*. Dalam penelitian ini hasil yang didapatkan menunjukkan bahwa metode LSTM tampil lebih baik jika dibandingkan dengan metode prediksi ARIMA dan *Back Propagation Neural Network* (BPNN). Menurut penulis metode *Long Short Term Memory Neural Network* (LSTM NN) merupakan metode yang pertama kali diterapkan dalam melakukan prediksi arus wisatawan [8].

2. Metode Penelitian

Dalam penelitian ini, variabel yang akan digunakan berjumlah tiga variabel yaitu *data event*, *data kalender*

pendidikan dan *data histori wisatawan*, ketiga variabel tersebut akan dijadikan variabel input *LSTM* dalam mempelajari pola untuk prediksi wisatawan. Data yang akan diujikan didapat dari data publik yaitu untuk *data event* didapat dari Dinas Kebudayaan dan Pariwisata Provinsi Nusa Tenggara Barat dari tahun 2017 sampai dengan tahun 2022, sedangkan untuk *dataset* kalender pendidikan didapat dari data publik kalender pendidikan dari tahun ajaran 2016/2017 sampai dengan tahun ajaran 2021/2022, dan untuk *data histori wisatawan* sebelumnya didapat dari website resmi NTB Satu Data dari tahun 2017 sampai dengan tahun 2022.

Hasil dari pengumpulan *dataset event*, kalender pendidikan dan *histori wisatawan* kemudian dilakukan konversi data menjadi data numerik dengan melakukan penghitungan jumlah *event*, liburan sekolah pada kalender pendidikan dan *histori jumlah wisatawan* yang terjadi pada setiap bulannya. Data yang telah dikonversi kedalam angka tersebut kemudian dimasukan kedalam tabel *dataset* seperti terlihat pada tabel 1 dibawah, dengan data variabel atau kolom yang terdiri dari variabel tanggal yang berisi data bulan untuk tiap tahun nya, variabel jumlah wisatawan yang berisi data jumlah pengunjung untuk setiap bulannya, variabel kalender *event* yang berisi jumlah *event* setiap bulan yang diadakan di Provinsi Nusa Tenggara Barat, variabel libur sekolah yang berisi jumlah liburan dari kalender pendidikan yang diambil untuk setiap bulannya. *Dataset* yang telah disusun dari ketiga variabel dapat dilihat pada tabel 1 berikut :

Tabel 1. *Dataset Event*, Libur Sekolah dan *Histori Jumlah Wisatawan*

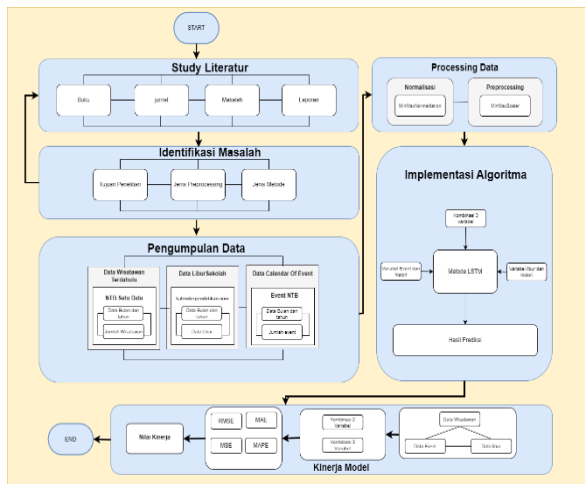
Tanggal	Jumlah Wisatawan	Kalender <i>event</i>	Libur Sekolah
01/01/2017	148.252	0	8
01/02/2017	162.279	1	0
01/03/2017	158.614	1	1
01/04/2017	221.701	3	2
01/05/2017	235.889	1	5
::	::	::	::
::	::	::	::
01/08/2022	127.042	6	1
01/09/2022	121.794	2	0
01/10/2022	125.265	4	1
01/11/2022	129.427	6	0
01/12/2022	154.209	4	15

Dari tabel 1 diatas data pada kolom 1 berisi data tanggal perbulan dalam tiap tahunnya, pada kolom kedua berisi data jumlah wisatawan yang merupakan histori data untuk tiap bulan pertahun nya dari tahun 2017 sampai dengan tahun 2022, pada kolom ketiga berisi data kalender *event* yaitu jumlah *event* yang terjadi pada setiap bulan, dan pada kolom yang terakhir berisi data liburan sekolah yang berisi jumlah liburan sekolah yang dihitung pada setiap bulannya.

2.1 Alur Penelitian

Tahapan penelitian yang dilakukan dalam pengujian data *event* dan liburan sekolah berdasarkan kalender pendidikan untuk melakukan prediksi terhadap jumlah

merupakan nilai maksimum dari setiap kolom dalam *matrix dataset* [13].



Pada gambar 1 diatas penelitian dimulai dari melakukan study literatur kemudian melakukan identifikasi masalah dan dilanjutkan ketahap pengumpulan data, setelah data berhasil dikumpulkan pada gambar 1 diatas dilanjutkan ketahap preprocessing data dengan menggunakan *MinMaxScaler Preprocessing*, tahapan ini dilakukan untuk melakukan normalisasi terhadap data yang terkadang data tersebut mengalami *missing value*, kesalahan pengukuran atau data yang ada mengalami *outlier* [9]. Setelah itu baru kemudian dilakukan implementasi pengujian *dataset* yang akan diuji menggunakan Algoritma yang dipilih yaitu *Long Short Term Memory (LSTM)*. Hasil dari pengujian algoritma kemudian dilakukan pengujian kinerja model dengan menggunakan *Mean Absolute Percentage Error (MAPE)*, *Mean Absolute Error (MAE)* dan *Root Mean Square Error (RMSE)*.

Long Short Term Memory (LSTM) merupakan salah satu metode hasil modifikasi dari *Recurrent Neural Network (RNN)*. Diantara hasil modifikasi RNN yang ada metode LSTM merupakan metode hasil modifikasi yang populer [14], LSTM dikembangkan untuk melengkapi kekurangan dari RNN yang tidak dapat melakukan prediksi berdasarkan informasi masa lalu yang disimpan dalam jangka waktu yang lama [15]. Dengan adanya LSTM kekurangan tersebut bisa dihindari, karena LSTM mampu mengingat kumpulan informasi yang telah disimpan dalam waktu yang lama, bahkan mampu menghapus informasi yang sudah tidak relevan lagi. Perbedaan yang paling utama antara RNN dan LSTM adalah pada LSTM memiliki empat lapisan modul jaringan dan informasi dipertukarkan dengan cara khusus [16]. Keunggulan *LSTM* lainnya adalah metode ini mampu mengontrol *overfitting* [17]

Dataset yang telah dimiliki harus dilakukan *preprocessing* agar nilai variabel dalam *dataset* menjadi rentang yang lebih terstandarisasi [10]. Dalam melakukan *pre-processing* data ada dua tahapan yang harus dilakukan yaitu tahapan pembersihan data dan tahapan normalisasi data [11]. Dalam melakukan normalisasi *dataset* pada penelitian ini menggunakan normalisasi *MinMaxScaler*, teknik ini merubah nilai yang ada menjadi nilai interval antara 0 dan 1 [12]. Adapun persamaan nya seperti pada persamaan 1 berikut.

Setelah pengujian model dengan menggunakan metode LSTM didapatkan hasilnya, selanjutnya adalah melakukan validasi terhadap model yang telah diuji dengan melihat seberapa besar tingkat *error* dari pengujian yang telah dilakukan adapun validasi pengujian menggunakan beberapa metode yaitu *root mean square error (RMSE)*, *mean absolute error (MAE)*, dan *mean absolute percentage error (MAPE)*. Adapun rumus perhitungan dari tiap pengujian kinerja dapat dilihat pada persamaan 2, persamaan3, dan persamaan 4 berikut.

Pada persamaan 1 diatas X_{scaled} merupakan variabel input sedangkan X merupakan *matrix dataset* yang akan dinormalisasi, X_{min} merupakan nilai minimum dari setiap kolom dalam *matrix dataset*, dan X_{max}

Pada persamaan 2 diatas n merupakan jumlah data, j merupakan indeks yang dimulai dari 1, y merupakan nilai aktual, dan y^1 merupakan nilai prediksi.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|et|}{y} \times 100 \quad (3)$$

Pada persamaan 3 diatas $\frac{1}{n}$ menunjukkan pembagian dengan jumlah total data atau observasi agar diperoleh rata-rata kesalahan mutlak, n adalah jumlah data, t merupakan indeks yang dimulai dari 1, $|et|$ Merupakan kesalahan absolut pada setiap observasi, dihitung dengan mengambil selisih antara nilai aktual dan nilai prediksi dan y merupakan nilai aktual.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (et)^2} \quad (4)$$

Pada persamaan 4 diatas $\frac{1}{n}$ menunjukkan pembagian dengan jumlah total data atau observasi agar diperoleh rata-rata kesalahan mutlak, n adalah jumlah data, t merupakan indeks yang dimulai dari 1, $|et|$ Merupakan kesalahan absolut pada setiap observasi.

3. Hasil dan Pembahasan

3.1 Pengolahan Data

Dataset yang telah dikumpulkan untuk penelitian terdiri dari variabel jumlah wisatawan perbulan, *event* dan juga libur sekolah. *Dataset* tersebut kemudian diunggah ke *folder content* pada *google colab* untuk bisa diakses oleh *google colab*.



tanggal	wisatawan	COE	libur
2017-01-01	148252	0	8
2017-02-01	162279	1	0
2017-03-01	158614	1	1
2017-04-01	221701	3	2
2017-05-01	235889	1	5
2017-06-01	306186	1	13
2017-07-01	422834	1	15

Gambar 2. *Dataset*

Pada gambar 2 diatas *dataset* yang tadinya berupa file *xlsx* kemudian dilakukan pembacaan (*load*) data dari *google colab* kemudian hasil dari pemanggilan data tersebut menghasilkan struktur data dengan *dataset* yang ada berisi variabel tanggal perbulan, hisotri jumlah wisatawan, *calender of event (COE)* yang ada di Provinsi Nusa Tenggara Barat dan kolom terakhir adalah variabel libur yang diambil dari kalender liburan sekolah.

3.2 Pengujian

Dalam melakukan perhitungan proses komputasi metode LSTM menggunakan beberapa bagian tahapan perhitungan, adapun persamaan *LSTM* adalah sebagai berikut :

$$i_t = \sigma_1(W_i x_t + U_i c_{t-1} + b_i) \quad (5)$$

Pada persamaan 5 diatas i_t merupakan *input gate*, σ merupakan *sigmoid*, W_i merupakan matrix bobot untuk *input gate*, x_t merupakan vektor input pada waktu, U_i merupakan matrix bobot untuk *cell stat*, c_{t-1} merupakan *cell state* pada waktu sebelumnya, dan b_i merupakan bias yang diberikan pada *input gate*.

$$f_t = \sigma_1(W_f x_t + U_f c_{t-1} + b_f) \quad (6)$$

Pada persamaan 6 diatas f_t merupakan *vector forget gate*, σ merupakan *sigmoid*, W_f merupakan matrix bobot untuk *input gate*, x_t merupakan vektor input pada waktu, U_f merupakan matrix bobot untuk *cell stat*, c_{t-1} merupakan *cell state* pada waktu sebelumnya, dan b_f merupakan bias yang diberikan pada *input gate*.

$$o_t = \sigma_1(W_o x_t + U_o c_{t-1} + b_o) \quad (7)$$

Pada persamaan 7 diatas o_t merupakan *output gate*, σ merupakan *sigmoid*, W_o merupakan matrix bobot untuk *input gate*, x_t merupakan vektor input pada waktu, U_o merupakan matrix bobot untuk *cell stat*, c_{t-1} merupakan *cell state* pada waktu sebelumnya, dan b_o merupakan bias yang diberikan pada *input gate*.

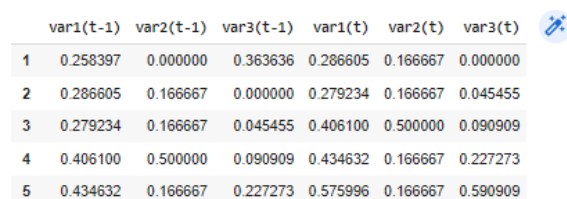
$$c_t = f_t * c_{t-1} + i * \sigma_1(W_c x_t + b_c) \quad (8)$$

Dari persamaan 8 diatas c_t merupakan *state sel memory*, f_t merupakan *vector forget gate*, i merupakan *input*, σ merupakan *sigmoid*, W_c merupakan matrix bobot untuk *input gate*, x_t merupakan vektor input pada waktu dan b_c merupakan vector bias.

$$h_t = o_t * \sigma_2(c_t) \quad (9)$$

Pada persamaan 9 diatas h_t merupakan *hidden state* waktu saat ini, o_t merupakan *output* dari *output gate*, σ_2 *sigmoid* untuk fungsi aktivasi tangen hiperbolik.

Selanjutnya pada tahap pengujian, struktur data yang telah *diload* oleh *google colab* Selanjutnya dilakukan normalisasi dengan menggunakan *MinMax-Scaler* supaya data siap digunakan dan dimasukan kedalam model. Dengan menggunakan rumus *MinMaxScaler* pada persamaan 1, data pada gambar 3 yaitu *var1*, *var2* dan *var3* yang tadinya masih berupa bilangan ribuan dikonversi menjadi nilai rentang 0 sampai 1, dari gambar 3 dibawah ini bisa dilihat hasil dari tiap variabel.



	var1(t-1)	var2(t-1)	var3(t-1)	var1(t)	var2(t)	var3(t)
1	0.258397	0.000000	0.363636	0.286605	0.166667	0.000000
2	0.286605	0.166667	0.000000	0.279234	0.166667	0.045455
3	0.279234	0.166667	0.045455	0.406100	0.500000	0.090909
4	0.406100	0.500000	0.090909	0.434632	0.166667	0.227273
5	0.434632	0.166667	0.227273	0.575996	0.166667	0.590909

Gambar 3. Hasil Normalisasi

Pada gambar 3 diatas data dinormalisasi dengan melakukan konversi menjadi rentang nilai antara 0 dan 1. Setelah data dilakukan normalisasi selanjutnya dilakukan pemilihan variabel tujuannya untuk menguji perbandingan nilai kinerja kesalahan menggunakan pengukuran *MAE*, *MAPE* dan *RMSE* dari tiap variabel

ataupun gabungan variabel itu sendiri. Adapun variabel yang akan digunakan dalam memprediksi dimulaidari pemilihan dan pengujian variabel histori wisatawan dan *event* kemudian dilakukan pengukuran kesalahannya, setelah itu variabel histori wisatawan dan libur sekolah sama juga seperti pengujian pada variabel sebelumnya, serta yang terakhir pengujian variabel histori wisatawan, *event* dan libur sekolah yang digunakan secara bersamaan sama juga dilakukan pengujian sampai tahap didapatkan nilai dari kesalahan menggunakan pengukuran *mae*, *mape* dan *rmse*. Pada gambar 4 dibawah ini adalah contoh pemilihan variabel kesatu (histori wisatawan) dan kedua (*event*) yang digunakan untuk pengujian

	var1(t-1)	var2(t-1)	var1(t)
1	0.258397	0.000000	0.286605
2	0.286605	0.166667	0.279234
3	0.279234	0.166667	0.406100
4	0.406100	0.500000	0.434632
5	0.434632	0.166667	0.575996

Gambar 4. Pemilihan Variabel Hisori Wisatawan dan *Event*

Dari gambar 4 diatas data yang digunakan dalam melakukan prediksi wisatawan hanya menggunakan variabel histori wisatawan yaitu *var1*, dan juga variabel *event* yaitu *var2*.

	var1(t-1)	var3(t-1)	var1(t)
1	0.258397	0.363636	0.286605
2	0.286605	0.000000	0.279234
3	0.279234	0.045455	0.406100
4	0.406100	0.090909	0.434632
5	0.434632	0.227273	0.575996

Gambar 5. Pemilihan Variabel Histori Wisatawan dan Liburan Sekolah

Dari gambar 5 diatas data yang digunakan dalam melakukan prediksi wisatawan hanya menggunakan variabel histori wisatawan yaitu *var1*, dan juga variabel liburan sekolah yaitu *var3*.

	var1(t-1)	var2(t-1)	var3(t-1)	var1(t)
1	0.258397	0.000000	0.363636	0.286605
2	0.286605	0.166667	0.000000	0.279234
3	0.279234	0.166667	0.045455	0.406100
4	0.406100	0.500000	0.090909	0.434632
5	0.434632	0.166667	0.227273	0.575996

Gambar 6 pemilihan Variabel Histori Wisatawan, *Event*, dan Libur Sekolah

Sedangkan pada gambar 6 diatas data yang digunakan dalam melakukan prediksi wisatawan menggunakan semua variabel yang digunakan yaitu histori wisatawan sebagai *var1*, variabel *event* sebagai *var2* dan variabel liburan sekolah sebagai *var3*. Ketiga pemilihan variabel tersebut digunakan untuk melihat kinerja pada masing-masing variabel pada *dataset* yang digunakan.

Tahapan selanjutnya setelah data dinormalisasi adalah melakukan pelatihan model, tahapan ini dilakukan dengan tujuan supaya bisa melihat hasil performa

kinerja yang dilakukan oleh metode LSTM, adapun hasil pelatihan model yang telah dilakukan adalah seperti berikut.

```
Epoch 196/200
2/2 - 0s - loss: 0.0510 - val_loss: 0.0078 - 28ms/epoch - 14ms/step
Epoch 197/200
2/2 - 0s - loss: 0.0509 - val_loss: 0.0077 - 26ms/epoch - 13ms/step
Epoch 198/200
2/2 - 0s - loss: 0.0508 - val_loss: 0.0075 - 26ms/epoch - 13ms/step
Epoch 199/200
2/2 - 0s - loss: 0.0507 - val_loss: 0.0074 - 25ms/epoch - 12ms/step
Epoch 200/200
2/2 - 0s - loss: 0.0507 - val_loss: 0.0073 - 30ms/epoch - 15ms/step
Epoch 1/200
2/2 - 2s - loss: 0.3034 - val_loss: 0.0587 - 2s/epoch - 884ms/step
Epoch 2/200
2/2 - 0s - loss: 0.2964 - val_loss: 0.0554 - 24ms/epoch - 12ms/step
Epoch 3/200
2/2 - 0s - loss: 0.2896 - val_loss: 0.0523 - 23ms/epoch - 11ms/step
Epoch 4/200
2/2 - 0s - loss: 0.2829 - val_loss: 0.0492 - 22ms/epoch - 11ms/step
Epoch 5/200
2/2 - 0s - loss: 0.2762 - val_loss: 0.0462 - 26ms/epoch - 13ms/step
```

Gambar 7. Melatih Model Neuron 20 Epoch 200 Batch Size 32

Dari gambar 7 bisa dilihat nilai *loss* dari epoch 200/200 menunjukkan nilai 0,2762 dan nilai *val_loss* adalah 0,0462 ini menunjukkan bahwa model dalam mempelajari *dataset* memiliki performa yang baik dalam mempelajari dan memprediksi data. Selanjutnya kita bandingkan dengan model dengan jumlah *neuron* 40, *epoch* 150 dan *batch size* 16 dapat dilihat pada gambar 8 dibawah.

```
Epoch 1/150
3/3 - 2s - loss: 0.3078 - val_loss: 0.0514 - 2s/epoch - 740ms/step
Epoch 2/150
3/3 - 0s - loss: 0.2976 - val_loss: 0.0480 - 24ms/epoch - 8ms/step
Epoch 3/150
3/3 - 0s - loss: 0.2877 - val_loss: 0.0447 - 24ms/epoch - 8ms/step
Epoch 4/150
3/3 - 0s - loss: 0.2780 - val_loss: 0.0416 - 26ms/epoch - 9ms/step
Epoch 5/150
3/3 - 0s - loss: 0.2686 - val_loss: 0.0386 - 23ms/epoch - 8ms/step
Epoch 146/150
3/3 - 0s - loss: 0.0507 - val_loss: 0.0129 - 26ms/epoch - 9ms/step
Epoch 147/150
3/3 - 0s - loss: 0.0507 - val_loss: 0.0129 - 31ms/epoch - 10ms/step
Epoch 148/150
3/3 - 0s - loss: 0.0506 - val_loss: 0.0128 - 30ms/epoch - 10ms/step
Epoch 149/150
3/3 - 0s - loss: 0.0505 - val_loss: 0.0128 - 30ms/epoch - 10ms/step
Epoch 150/150
3/3 - 0s - loss: 0.0505 - val_loss: 0.0127 - 32ms/epoch - 11ms/step
```

Gambar 8. Melatih Model Neuron 40 Epoch 200 Batch Size 32

Pada gambar 8 diatas hasil *loss* pada *epoch* terakhir memiliki nilai lebih kecil dibanding dengan model pertama dan nilai *val loss* juga menghasilkan nilai yang lebih kecil sehingga model ini memiliki model lebih baik. Untuk membuktikan model terbaik digunakan metode pencarian hyperparameter dalam menemukan kombinasi terbaik untuk menghasilkan kinerja model yang optimal. Pada gambar 9 dibawah dapat dilihat hyperparameter terbaik untuk menghasilkan kinerja model.

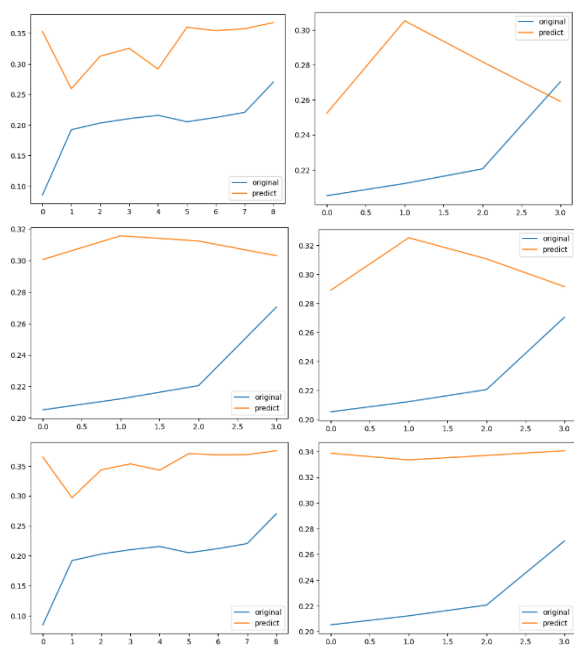
```
2/2 [=====] - 0s 4ms/step - loss: 0.0494
Epoch 177/200
2/2 [=====] - 0s 4ms/step - loss: 0.0480
4
print('Best hyperparameters: ', random_search.best_params_)
Best hyperparameters: {'num_neurons': 40, 'epochs': 200, 'dropout': 0.5, 'batch_size': 16}
```

Gambar 9. Metode *RandomizedSearch* Dalam Mencari Hyperparameter Terbaik

Setelah model dilatih tahapan selanjutnya adalah pengujian variable, Dalam pengujian variabel *event* dan

libur kalender pendidikan untuk melakukan prediksi terhadap jumlah kedatangan wisatawan ke Provinsi Nusa Tenggara Barat dilakukan beberapa kali pengujian dengan menguji variabel independen yaitu variabel *event* dan libur sekolah serta variabel dependen yaitu jumlah wisatawan.

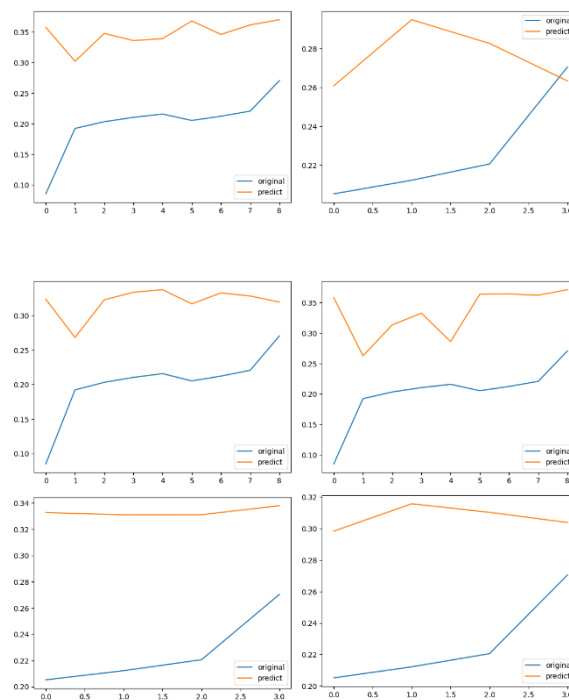
Untuk pengujian prediksi tiap variabel yaitu variabel histori, *event* dan liburan, histori dan *event* saja, serta histori dan liburan saja disajikan berdasarkan dua parameter pengujian seperti terlihat pada gambar grafik berikut yaitu gambar 10 pengujian dengan parameter 1 dan gambar 11 merupakan pengujian dengan parameter 2.



Gambar 10. Prediksi Dengan Parameter *Neuron : 40, Epoch : 150, Batch_size : 16*

Pada gambar 10 diatas untuk bagian gambar yang paling atas hasil pengujian semua variabel dengan *data split* 80/20 untuk yang kiri, dan untuk yang kanan menggunakan *data split* 90/10. Sedangkan pada bagian yang tengah yaitu gambar kedua pada hasil gambar yang kiri adalah hasil pengujian variabel *event* dengan *data split* 80/20 sedangkan bagian kanan pengujian variabel *event* dengan *data split* 90/10. Pada gambar paling bawah untuk bagian kanan adalah hasil pengujian untuk variabel libur sekolah dengan pembagian *data split* 80/20 dan untuk bagian kanan dengan pembagian *data split* 90/10.

Adapun untuk pengujian dengan menggunakan parameter kedua dengan menggunakan parameter *neuron : 20, epoch : 200*, dan *batch_size : 32*, hasil pengujian dapat dilihat seperti pada gambar 11 dibawah ini.



Gambar 11. Prediksi Dengan Parameter *Neuron : 20, Epoch : 200, Batch_size : 32*

Setelah itu kemudian dilakukan pengujian untuk melihat hasil prediksi dan aktual, hasil pengujian dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Hasil Prediksi Parameter 1 *splitting* 90/10

Bulan	Variabel	Prediksi	Aktual
2022-08-01	semua	160180.12	127042.01
2022-09-01	Semua	150496.44	121794
2022-10-01	Semua	175510.52	125265.01
2022-11-01	Semua	165454.45	129427
2022-12-01	Semua	153400.08	154209.02
2022-08-01	Liburan	164489.05	127042.01
2022-09-01	Liburan	181958.03	121794
2022-10-01	Liburan	180499.02	125265.01
2022-11-01	Liburan	181129.1	129427
2022-12-01	Liburan	184060.	154209.02
2022-08-01	Event	172533.22	127042.01
2022-09-01	Event	155672.56	121794
2022-10-01	Event	169616.19	125265.01
2022-11-01	Event	165008.25	129427
2022-12-01	Event	156856.47	154209.02

Data pada tabel 2 diatas menunjukkan hasil pengujian prediksi wisatawan 5 bulan terakhir dengan menggunakan *Neuron : 20, Epoch : 200, Batch_size : 32* dengan *splitting dataset* 90/10.

Hasil prediksi dengan menggunakan parameter yang sama dan dengan *splitting dataset* 80/20 dapat dilihat pada tabel 3 berikut.

Tabel 3. Hasil prediksi Parameter 1 *splitting* 80/20

Bulan	Variabel	Prediksi	Aktual
2022-08-01	Semua	144938.34	127042.01
2022-09-01	Semua	194207.97	121794
2022-10-01	Semua	192103.48	125265.01
2022-11-01	Semua	193268.92	129427
2022-12-01	Semua	198750.64	154209.02
2022-08-01	Liburan	161003.53	127042.01
2022-09-01	Liburan	196712.92	121794

2022-10-01	Liburan	196552.06	125265.01
2022-11-01	Liburan	195847.23	129427
2022-12-01	Liburan	200293.16	154209.02
2022-08-01	Event	192220.53	127042.01
2022-09-01	Event	197312.25	121794
2022-10-01	Event	194311.39	125265.01
2022-11-01	Event	198427.17	129427
2022-12-01	Event	198344.6	154209.02

Adapun hasil prediksi dengan menggunakan parameter lain yaitu jumlah *neuron* 40, *epoch* 150 dan *batch size* 16 bisa dilihat pada tabel 4 dan tabel 5 dibawah ini.

Tabel 4. Hasil prediksi Parameter 2 *splitting* 90/10

Bulan	Variabel	Prediksi	Aktual
2022-09-01	Semua	150789.4	121794
2022-10-01	Semua	177136.02	125265.01
2022-11-01	Semua	166674.64	129427
2022-12-01	Semua	153963.38	154209.02
2022-09-01	Liburan	183097.03	121794
2022-10-01	Liburan	180402.39	125265.01
2022-11-01	Liburan	182205.08	129427
2022-12-01	Liburan	184229.61	154209.02
2022-09-01	Event	152726.36	121794
2022-10-01	Event	176348.55	125265.01
2022-11-01	Event	166807.48	129427
2022-12-01	Event	153946.97	154209.02

Dan terakhir parameter 2 dengan *splitting* 80/20 dapat dilihat pada tabel 5 berikut.

Tabel 5. Hasil prediksi Parameter 2 *splitting* 80/20

Bulan	Variabel	Prediksi	Aktual
2022-08-01	Semua	184150.5	127042.01
2022-09-01	Semua	195997.52	121794
2022-10-01	Semua	197384.86	125265.01
2022-11-01	Semua	198216.25	129427
2022-12-01	Semua	198187.	154209.02
2022-08-01	Liburan	155651.08	127042.01
2022-09-01	Liburan	194614.52	121794
2022-10-01	Liburan	194273.39	125265.01
2022-11-01	Liburan	193681.	129427
2022-12-01	Liburan	198289.4	154209.02
2022-08-01	Event	193871.14	127042.01
2022-09-01	Event	178293.1	121794
2022-10-01	Event	192476.98	125265.01
2022-11-01	Event	188946.81	129427
2022-12-01	Event	179531.12	154209.02

Setelah hasil prediksi dihasilkan kemudian dilakukan pengujian kinerja untuk mengukur tingkat akurasi dari model yang dibuat, dalam melakukan pengukuran kinerja digunakan tiga pengukuran yaitu MAPE, RMSE dan MAE dengan perhitungan seperti pada persamaan 2, persamaan 3 dan persamaan 4 . Hasil pengukuran dapat dilihat pada tabel 6 berikut :

Tabel 6. Pengujian dengan parameter *Neuron* : 40, *Epoch* :150, *Batch_size* : 16

No	Variabel	Splitting	MAPE	RMSE	MAE
1	Histori & Libur	90/10	38.65	51198.03	49809.77
2	Histori & Event	90/10	23.81	35226.79	29914.60
3	Histori, Even & Liburan	90/10	20.84	30211.36	26445.22
4	Histori & Libur	80/10	72.7	79538.18	75682.73
5	Histori & Event	80/10	62.84	69177.28	64527.26

6	Histori, Even & Liburan	80/10	67.63	74121.67	69817.51
---	-------------------------	-------	-------	----------	----------

Pada tabel 6 diatas nilai MAPE terkecil didapatkan pada pengujian semua variabel yaitu wisatawan, *event* dan liburan dengan nilai MAPE sebesar 20,84 dengan *data training* 90% dan *data testing* 10%. Sedangkan pada tabel 7 dibawah menggunakan jumlah *neuron* 20, *epoch* 200 dan *batch size* 32.

Tabel 7. Pengujian parameter *neuron*: 20, *epoch*:200, *batch_size*: 32

No	Variabel	Split	MAPE	RMSE	MAE
1	Histori, Even & Liburan	90/10	23	33897.5	29784.5
2	Histori & libur	90/10	36.17	48234.2	46551.4
3	Histori & Event	90/10	25.6	35936.8	32389.9
4	Histori, Even & Liburan	80/10	55.93	64337.2	56060.5
5	Histori & libur	80/10	61.68	68906.8	62841.2
6	Histori & Event	80/10	69.9	76253.1	72496.2

Berdasarkan hasil pengujian dengan menggunakan berbagai parameter didapatkan hasil terbaik menggunakan data *splitting* 90% untuk *training* dan 10% untuk *testing*, dengan parameter terbaik jumlah *neuron* 40, jumlah *epoch* 150 dan *batch size* 16. Hal ini sesuai dengan hasil pencarian hyperparameter terbaik menggunakan *random size search*.

4. Kesimpulan

Dalam penelitian yang dilakukan terhadap variabel independen *event* dan libur sekolah terhadap variabel dependen wisatawan menggunakan metode LSTM maka dapat disimpulkan, *dataset* yang digunakan dari tahun 2017 sampai dengan 2022, akan tetapi dikarenakan pada tahun 2020 dan 2021 mengalami Covid-19 sehingga pemerintah setempat menerapkan Pembatasan Sosial Berskala Besar (PSBB) maka *dataset* pada tahun tersebut tidak digunakan untuk menghindari *overfitting* pada model.

Pengujian menggunakan *dataset* sebanyak 47 data dengan melakukan berbagai pengujian dengan kombinasi variabel yang berbeda, serta penggunaan 2 parameter yang berbeda dan juga *splitting* yang berbeda, didapatkan nilai terbaik pada pengujian gabungan variabel histori, *event*, dan liburan *splitting* 90/10 pada parameter *neuron* : 40, *epoch* :150, *batch_size* : 16 dengan nilai kinerja MAE 26445.22 MAPE 20.84%, dan RMSE 30211.36. Dengan hasil ini menunjukan bahwa LSTM dapat meningkatkan akurasi pada prediksi dan memberikan nilai prediksi yang lebih mendekati dengan nilai aktual.

Penelitian ini menunjukan bahwa LSTM dapat mempelajari hubungan kompleks antara variabel histori wisatawan, variabel *event* dan liburan sekolah sehingga LSTM dapat mengintegrasikan variabel-

variabel tersebut kedalam model prediksi dengan sangat baik.

Untuk mendapatkan tingkat kinerja *MAE*, *MAPE*, *RMSE* yang lebih kecil diharapkan pada penelitian berikutnya *dataset* yang digunakan bisa lebih banyak lagi, dan bisa menggunakan metode *non linear* lain supaya mendapatkan tingkat *error* yang lebih kecil dalam melakukan prediksi wisatawan.

Daftar Rujukan

- [1] N. P. N. Hendayanti, I. K. P. Suniantara, and M. Nurhidayati, "Penerapan Support Vector Regression (Svr) Dalam Memprediksi Jumlah Kunjungan Wisatawan Domestik Ke Bali," *Jurnal Varian*, vol. 3, no. 1, pp. 43–50, 2019.
- [2] 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.
- [3] M. A. Ridla, N. Azise, and M. Rahman, "Perbandingan Model Time Series Forecasting Dalam Memprediksi Jumlah Kedatangan Wisatawan Dan Penumpang Airport," *Jurnal Sistem Informasi dan Sistem Komputer*, vol. 8, no. 1, pp. 1–14, 2023.
- [4] K. T. N. Lestari, M. A. Albar, and R. Afwani, "Penerapan Metode Backpropagation Dalam Memprediksi Jumlah Kunjungan Wisatawan Ke Provinsi Nusa Tenggara Barat (NTB)," *Journal of Computer Science and Informatics Engineering (J-Cosine)*, vol. 3, no. 1, pp. 39–48, 2019.
- [5] R. Rahmawati, Y. Yuniza, A. N. Rahma, and Z. Zukrianto, "PREDIKSI JUMLAH WISATAWAN DI KOTA PEKANBARU PADA TAHUN 2019-2023 DENGAN MENGGUNAKAN METODE FUZZY TIME SERIES CHEN," *THETA: Jurnal Pendidikan Matematika*, vol. 2, no. 1, pp. 36–44, 2020.
- [6] J.-W. Bi, Y. Liu, and H. Li, "Daily tourism volume forecasting for tourist attractions," *Ann Tour Res*, vol. 83, p. 102923, 2020.
- [7] H. Mukhtar, Y. Rizki, F. A. Wenando, and M. A. Al Aziz, "Prediksi Kunjungan Wisatawan dengan Reduksi Noise pada Google Trends menggunakan Hilbert-Huang Transform dan Long Short-Term Memory," *JURNAL FASILKOM*, vol. 12, no. 3, pp. 152–159, 2022.
- [8] Y. Li and H. Cao, "Prediction for tourism flow based on LSTM neural network," *Procedia Comput Sci*, vol. 129, pp. 277–283, 2018.
- [9] V. R. Prasetyo, B. Hartanto, and A. A. Mulyono, "Penentuan pembimbing tugas akhir mahasiswa jurusan teknik informatika universitas surabaya dengan metode dice coefficient," *Teknika*, vol. 8, no. 1, pp. 44–51, 2019.
- [10] K. Dewi, P. P. Adikara, and S. Adinugroho, "Prediksi Indeks Harga Konsumen (IHK) Kelompok Perumahan, Air, Listrik, Gas Dan Bahan Bakar Menggunakan Metode Support Vector Regression," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, vol. 2548, p. 964X, 2018.
- [11] R. Firdaus and H. Mukhtar, "Prediksi Indeks Harga Produsen Pertanian Karet Di Indonesia Menggunakan Metode LSTM," *JURNAL FASILKOM*, vol. 13, no. 01, pp. 1–6, 2023.
- [12] D. Heksaputra, "Prediksi Rerata Harga Beras Tingkat Grosir Indonesia dengan Long Short Term Memory," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 2, pp. 163–174, 2020.
- [13] H. Mukhtar, R. Gunawan, A. Hariyanto, and W. Mulyana, "Peramalan Kedatangan Wisatawan ke Suatu Negara Menggunakan Metode Support Vector Machine (SVM)," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 274–282, 2022.
- [14] R. F. Hidayat, B. Irawan, and M. Nasrun, "Implementasi Klasifikasi Gambar Untuk Industri Pakaian Menggunakan Image Search Engine Berbasis Website," *eProceedings of Engineering*, vol. 10, no. 1, 2023.
- [15] J. Manullang, A. J. Santoso, and A. W. R. Emanuel, "Prediksi Kunjungan Wisatawan Taman Nasional Gunung Merbabu dengan Time Series Forecasting dan LSTM," *Jurnal Teknik Sipil*, vol. 11, no. 2, pp. 132–140, 2020.
- [16] A. Nilsen, "Perbandingan Model RNN, Model LSTM, dan Model GRU dalam Memprediksi Harga Saham-Saham LQ45," *Jurnal Statistika dan Aplikasinya*, vol. 6, no. 1, pp. 137–147, 2022.
- [17] M. Rizki, S. Basuki, and Y. Azhar, "Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory (LSTM) Untuk Prediksi Curah Hujan Kota Malang," *Jurnal Repositor*, vol. 2, no. 3, pp. 331–338, 2020.
- [18] Y. P. Adiwiguna, H. Purbaseno, H. Syuhara, F. J. Pribadi, A. Sarbani, and A. D. Kusuma, "MEMBANGUN TOOL PREDIKSI YANG KOMPREHENSIF GUNA MENINGKATKAN AKURASI PRODUK AKHIR PENCAAMPURAN KUALITAS BATUBARA MENERAPKAN PERBANDINGAN KOLABORATIF PADA ALGORITMA REGRESI HINGGA NEURAL NET," *Indonesian Mining Professionals Journal*, vol. 4, no. 2, pp. 111–122, 2022.
- [19] S. Lionetti, D. Pfäffli, M. Pouly, T. vor der Brück, and P. Wegelin, "Tourism Forecast with Weather, Event, and Cross-industry Data," in *ICAART (2)*, 2021, pp. 1097–1104.
- [20] N. A. Jauhariyah, H. Habibulloh, and A. A. Yazid, "Potret Pertumbuhan Kunjungan Wisatawan Dalam Perspektif Islam Di Kabupaten BANYUWANGI," *Jurnal Ekonomi Syariah Darussalam*, vol. 2, no. 1, pp. 86–99, 2021.