

TESIS

**PENGUJIAN VARIABEL EVENT TAHUNAN DAN LIBUR SEKOLAH
DALAM MEMPREDIKSI JUMLAH WISATAWAN
(Studi Kasus: Provinsi Nusa Tenggara Barat)**



Disusun oleh:

Nama : Candra Rusmana
NIM : 22.55.1194
Konsentrasi : Digital Transformation Intellegent

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

TESIS

PENGUJIAN VARIABEL EVENT TAHUNAN DAN LIBUR SEKOLAH DALAM MEMPREDIKSI JUMLAH WISATAWAN (Studi Kasus: Provinsi Nusa Tenggara Barat)

TESTING THE VARIABLES OF ANNUAL EVENTS AND SCHOOL HOLIDAYS IN PREDICTING TOURIST NUMBERS (Case Study: West Nusa Tenggara Province)

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Candra Rusmana
NIM : 22.55.1194
Konsentrasi : Digital Transformation Intellegent

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PENGESAHAN

**PENGUJIAN VARIABEL EVENT TAHUNAN DAN LIBUR SEKOLAH DALAM
MEMPREDIKSI JUMLAH WISATAWAN
(Studi Kasus: Provinsi Nusa Tenggara Barat)**

**TESTING THE VARIABLES OF ANNUAL EVENTS AND SCHOOL HOLIDAYS IN
PREDICTING TOURIST NUMBERS.**

Dipersiapkan dan Disusun oleh

Candra Rusmana

22.55.1194

Telah Ditujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Kamis, 04 Januari 2024

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 04 Januari 2024

Rektor

Prof. Dr. M. Suvanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

**PENGUJIAN VARIABEL EVENT TAHUNAN DAN LIBUR SEKOLAH DALAM
MEMPREDIKSI JUMLAH WISATAWAN
(Studi Kasus: Provinsi Nusa Tenggara Barat)**

**TESTING THE VARIABLES OF ANNUAL EVENTS AND SCHOOL HOLIDAYS IN
PREDICTING TOURIST NUMBERS.**

Dipersiapkan dan Disusun oleh

Candra Rusmana

22.55.1194

Telah Ditujakan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Kamis, 04 Januari 2024

Pembimbing Utama

Anggota Tim Penguji

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

Tonny Hidayat, M.Kom., Ph.D.
NIK. 190302182

Pembimbing Pendamping

Hanif Al Fatta, M.Kom., Ph.D.
NIK. 190302096

Kusnawi, M.Kom.
NIK. 190302112

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 4 Januari 2024
Direktur Program Pascasarjana

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Candra Rusmana
NIM : 22.55.1194
Konsentrasi : Digital Transformation Intelligence

Menyatakan bahwa Tesis dengan judul berikut:
**PENGUJIAN VARIABEL EVENT TAHUNAN DAN LIBUR SEKOLAH
DALAM MEMPREDIKSI JUMLAH WISATAWAN (Studi Kasus: Provinsi
Nusa Tenggara Barat)**

Dosen Pembimbing Utama : Prof. Kusrini, M.Kom
Dosen Pembimbing Pendamping : Kusnawi, M.Kom

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta **sanksi lainnya** sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 04 Januari 2024
Yang Menyatakan,



Candra Rusmana

HALAMAN PERSEMBAHAN

Alhamdulillah puji syukur kehadiran Allah SWT yang telah memberikan kemudahan dan kelancaran kepada saya sehingga laporan tesis ini dapat terselesaikan. Laporan Tesis ini saya persembahkan untuk :

1. Kedua Orang Tua tercinta (Ibu Yaroh dan Bapak Apep Pendi) dan juga keluarga yang telah memberikan semangat dan juga do'a, semoga Allah memberikan Balasan terbaik bagi keduanya.
2. Istriku tercinta (Evi Sulastri), anak-anakku yang paling tersayang (Latifa Nida Shidqia yang ada di Syurga, Shabrina Nada Shidqia, Azzeema Auliani Zakiyya, dan M. Hijra Karim Assyakir) yang tidak henti-hentinya memberikan dukungan, dorongan, serta selalu menjadi motivasi yang kuat dalam melakukan segala hal. Semoga selalu dikelilingi rahmat, keberkahan dan kebahagiaan dunia dan akhirat.
3. Ibu Prof. Dr. Kusrini, M.Kom dan Bapak Kusnawi, M.Kom yang telah memberikan bimbingan aktif selama pelaksanaan penelitian, semoga selalu diberikan kemudahan dan keberkahan dalam segala urusannya.
4. Keluarga Besar, teman-teman seperjuangan serta semua pihak yang tidak bisa disebutkan satu persatu, yang selalu memberikan dukungan dan semangat dalam segala hal semoga segala kebaikan kalian dibalasa dengan balasan terbaik.

HALAMAN MOTTO

Jangan takut melangkah dalam mencari ilmu, karena Allah pasti akan memberikan pertolongan dan kemudahan bagi setiap orang yang ingin mencari ilmu, walaupun pasti ada saja halangan dan rintangan ketika mulai melangkah.

(“ Saya percaya setiap manusia memiliki jumlah detak jantung yang terbatas. Saya tidak bermaksud menyia-nyitakan milik saya. ” — Neil Armstrong)



KATA PENGANTAR

Puji Syukur senantiasa kehadiran Illaahi Rabbi yang telah memberikan karunia, hidayah, berkah serta kemudahan bagi penulis sehingga tesis ini dapat terselesaikan.

Tesis ini disusun dalam rangka memenuhi salah satu syarat kelulusan perguruan tinggi Program Studi Strata-2 Magister Teknik Informatika di Universitas Amikom Yogyakarta dan untuk meraih gelar Magister Komputer (M.Kom). selain dari itu tesis ini juga bertujuan untuk menggali pengetahuan tentang penggunaan variabel baru yaitu variabel event dan liburan sekolah berdasarkan kalender akademik dalam melakukan prediksi wisatawan.

Dalam kelancaran penyusunan tesis ini, penulis mengucapkan banyak terima kasih kepada :

1. Bapak Prof. Dr. Suyanto, MM selaku Rektor Amikom Yogyakarta.
2. Ibu Prof. Kusriani, M.Kom dan Bapak Kusnawi, M.Kom selaku dosen pembimbing yang telah memberikan masukan, saran, bantuan dan bimbingan langsung dalam menyelesaikan naskah tesis ini.
3. Ibu Prof. Kusriani, M.Kom selaku direktur Program Pasca Sarjana Universitas Amikom Yogyakarta.
4. Kedua orang tua yang selalu memberikan do'a dan support.
5. Istri dan anak-anakku yang selalu menjadi penyemangat.
6. Dosen Amikom Yogyakarta yang telah mencurahkan Ilmu dan pengalaman yang bermanfaat

Semua pihak yang telah membantu dalam menyelesaikan tesis ini yang tidak bisa disebutkan satu persatu. Mohon maaf yang sebesar-besarnya dari penulis apabila dalam penyusunan tesis ini masih banyak kekurangan dan masih belum mencapai kata sempurna. Saran dan kritik penulis harapkan dari semua pembaca sehingga dikemudian hari bisa menjadi suatu hal yang lebih baik.

Semoga tesis ini dapat menjadi bahan untuk menambah pengetahuan serta bisa memberikan manfaat bagi pembacanya khususnya bagi penulis sendiri. Semoga tesis ini bisa menjadi salah satu referensi untuk penelitian dikemudian hari.

Mataram, 04 Januari 2024

Penulis



DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	9
1.3. Batasan Masalah.....	9
1.4. Tujuan Penelitian.....	10
1.5. Manfaat Penelitian.....	10
BAB II TINJAUAN PUSTAKA.....	12
2.1. Tinjauan Pustaka.....	12
2.2. Keaslian Penelitian.....	20
2.3. Landasan Teori.....	27
BAB III METODE PENELITIAN.....	35

3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	35
3.2. Metode Pengumpulan Data.....	35
3.3. Metode Analisis Data.....	37
3.4. Alur Penelitian	38
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	41
4.1. Membangun Dataset	41
4.1.1. Pengumpulan Data	41
4.1.2. Pengujian Dataset.....	43
4.1.3. <i>Preprocessing Data</i>	44
4.2. Membangun Model.....	48
4.2.1 <i>Training dataset</i>	49
4.2.2 <i>Pengujian Metode LSTM</i>	52
4.2.3 <i>Pengujian Metode Support Vector Regression (SVR)</i>	60
BAB V PENUTUP.....	67
5.1. Kesimpulan	67
5.2. Saran	68
DAFTAR PUSTAKA	70

DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan posisi penelitian.....	20
Tabel 2.1. Lanjutan	21
Tabel 2.1. Lanjutan	22
Tabel 2.1. Lanjutan	23
Tabel 2.1. Lanjutan	24
Tabel 2.1. Lanjutan	25
Tabel 2.1. Lanjutan	26
Tabel 4.1. Hasil prediksi Parameter 1 semua variabel.....	54
Tabel 4.2. Hasil pengujian variabel libur parameter 1	55
Tabel 4.3 pengujian variabel <i>event</i> dengan parameter 1	55
Tabel 4.4. perbandingan nilai pengujian	56
Tabel 4.5 Hasil prediksi parameter 2 semua variabel	56
Tabel 4.6 Hasil pengujian variabel libur sekolah parameter 2.....	57
Tabel 4.7. Hasil pengujian variabel <i>event</i> Parameter 2.....	57
Tabel 4.8. perbandingan nilai pengujian parameter 2.....	58
Tabel 4.9. Pengujian parameter <i>Neuron</i> : 40, <i>Epoch</i> :150, <i>Batch_size</i> : 16.....	59
Tabel 4.10. Pengujian parameter <i>Neuron</i> : 20, <i>Epoch</i> :200, <i>Batch_size</i> : 32	59
Tabel 4.11. Pengujian <i>splitting</i> 70/30	61
Tabel 4.12 pengujian <i>splitting</i> 80/20.....	63
Tabel 4.13. Pengujian <i>splitting</i> 90/10	64
Tabel 4.14. Hasil pengujian kinerja	65

DAFTAR GAMBAR

Gambar 1.1. Jumlah wisatawan pertahun Prov. NTB	1
Gambar 2.1 Kondisi Sel Modul LSTM	29
Gambar 3.1. Metode analisis pengolahan data untuk prediksi.....	38
Gambar 3.2 Mind Map penelitian	39
Gambar 3.3 Alur Penelitian.....	40
Gambar 4.1. dataset yang telah diupload kedalam google colab.....	44
Gambar 4.2 hasil <i>preprocessing minmax scaller</i>	45
Gambar 4.3 pemilihan variabel metode <i>LSTM</i> var 1 dan 2.....	45
Gambar 4.4 pemilihan variabel metode <i>LSTM</i> variabel 1 dan 3.....	46
Gambar 4.5 pemilihan variabel metode <i>LSTM</i> semua variabel	46
Gambar 4.6 pemilihan variabel metode <i>SVR</i> semua variabel	47
Gambar 4.7 pemilihan variabel metode <i>SVR</i> wisatawan dan libur	48
Gambar 4.8 pemilihan var metode <i>SVR</i> wisatawan dan event	49
Gambar 4.9. Melatih Model Neuron 20 Epoch 200 Batch Size 32.....	50
Gambar 4.10. Melatih Model Neuron 40 Epoch 200 Batch Size 32	51
Gambar 4.14 hasil pencarian parameter terbaik <i>SVR</i>	52
Gambar 4.15. Prediksi parameter Neuron : 40, Epoch :150, Batch_size : 16.....	53
Gambar 4.16. Prediksi parameter Neuron : 20, Epoch :200, Batch_size : 32.....	54
Gambar 4.17. Pengujian <i>splitting</i> 70/30.....	60
Gambar 4.18 pengujian <i>SVR</i> dengan data <i>splitting</i> 80/20	62
Gambar 4.19. Prediksi <i>splitting</i> 90/10.....	63

INTISARI

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 didapatkan 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 *Long Short Term Memory (LSTM)* dan *Support Vector Regression (SVR)* dengan menggunakan variabel histori wisatawan, *event* dan libur sekolah. Pengujian LSTM dengan penggunaan variabel campuran menghasilkan kinerja terbaik MAPE sebesar 36.25% dengan penggunaan data training dan data testing 80/20. Hasil kinerja dengan variabel liburan saja menghasilkan kinerja terbaik MAPE sebesar 30.343%. Sedangkan hasil dengan variabel *event* saja menghasilkan kinerja terbaik MAPE sebesar 38.981%. Pengujian SVR dengan menggunakan variabel campuran menghasilkan kinerja terbaik MAPE 41.3125%, pengujian *event* saja menghasilkan kinerja terbaik 41.3125%, pengujian liburan menghasilkan kinerja terbaik 41.3124% Ini menunjukan bahwa variabel *event* dan libur sekolah bisa dengan baik digunakan dalam melakukan prediksi terhadap kedatangan wisatawan.

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

ABSTRACT

The event held in various regions is a special attraction for tourists to come to the place. School seasonal holidays are also an annual family agenda to go to tourist attractions. The ups and downs of the number of tourists coming to NTB Province have an impact on the local government, the community around tourist attractions and business actors in the field of tourism.

The purpose of this study is to test the annual event variables and school holidays. The dataset used was obtained from the NTB Province Public Website, namely data.ntbprov.go.id, the dataset was in the form of a history of the number of tourist visits every month, from the website disbudpar.ntbprov.go.id, the annual event was obtained and from the calendar website. For school holidays, dataset taken from each source is taken from 2017 to 2022. Of all the datasets obtained can be utilized in digging information to make predictions.

In predicting the Long Short Term Memory (LSTM) algorithm and Support Vector Regression (SVR) are used by using tourist history variables, events and school holidays. LSTM testing with the use of mixed variables produces the best performance of Mape of 36.25% with the use of training data and 80/20 testing data. Performance results with vacation variables alone produce the best performance of Mape of 30.343%. While the results with the event variable only produced the best performance of Mape by 38.981%. SVR testing using a mixed variable produces the best performance of Mape 41.3125%, Event testing alone produces the best performance of 41.3125%, holiday testing produces the best performance of 41.3124% shows that the event variable and school holidays can be well used in predicting tourist arrivals.

Keyword: event, holidays, LSTM, prediction, SVR, tourists

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Sebagai pulau yang dikelilingi oleh banyak pantai tidak heran pulau lombok memiliki pantai-pantai yang sangat indah. Dengan air pantainya yang sangat jernih dan ombaknya yang tidak terlalu besar sehingga *IndonesianTrivcanvas.co* menyebut bahwa Pulau Lombok sebagai saudara kembar dari Pulau Bali. Peranan sektor pariwisata ini bisa menjadikan peningkatan terhadap pendapatan ekonomi masyarakat sekitar Provinsi Nusa Tenggara Barat. Dapat dikatakan pula bahwa semakin tinggi populasi wisatawan yang datang kesuatu daerah maka akan semakin positif pula pengaruh terhadap peningkatan ekonomi masyarakat sekitar (Rizal dan Hartati, 2017).

Jumlah wisatawan dipulau lombok mengalami pasang surut setiap tahunnya. hal ini dapat dilihat dari grafik data wisatawan yang diambil dari website resmi Dinas Kebudayaan dan Pariwisata Provinsi Nusa Tenggara Barat berikut



Gambar 1.1. Jumlah wisatawan pertahun Prov. NTB

Jumlah kunjungan wisatawan yang datang ke Provinsi Nusa Tenggara Barat menurut gambar 1 diatas mengalami kenaikan dan juga penurunan. Ketika penyusun mengambil data dari website resmi Dinas Kebudayaan dan Pariwisata Provinsi Nusa Tenggara Barat pada laman <http://www.disbudpar.ntbprov.go.id/> data yang ditampilkan hanya sampai dengan tahun 2021. Pada gambar 1 terlihat jelas di tahun 2017 sampai dengan tahun 2018 mengalami penurunan dari total 3.508.903 pada menjadi 2.812.379 orang. Sedangkan pada tahun 2019 mengalami lagi peningkatan menjadi 3.706.352 orang dan pada tahun 2019 merupakan angka tertinggi dalam jumlah wisatawan yang datang ke pulau lombok. Kemudian pada tahun 2020 terjadi penurunan yang sangat signifikan karena jumlah wisatawan hanya berjumlah 400.595 orang, dan pada tahun 2021 mengalami penurunan kembali dengan jumlah wisatawan hanya 344.733 orang. Sedangkan data terbaru yang didapat dari data publik NTB satu data, pada tahun 2022 jumlah wisatawan mengalami kenaikan dengan jumlah 1.121.436 orang. Hal ini tidak sesuai dengan target dari pemerintah Provinsi Nusa Tenggara Barat yang menargetkan pada tahun 2022 jumlah wisatawan yang datang adalah 4,5 juta. Sedangkan untuk target tahun 2023 Pemerintah Provinsi Nusa Tenggara Barat menargetkan jumlah wisatawan yang datang ke Provinsi NTB adalah sebanyak 2,5 juta orang. Target pemerintah tersebut akan bisa lebih *realistis* jika pemerintah mengungkapkan target tersebut berdasarkan data-data yang disajikan, supaya target pemerintah tidak terlalu jauh dengan data sebenarnya, maka dari itu dibutuhkan prediksi yang akurat dalam membantu memprediksi target pemerintah tersebut.

Jumlah kedatangan wisatawan ke Provinsi Nusa Tenggara Barat baik wisatawan mancanegara ataupun wisatawan lokal memiliki berbagai waktu dan momen, ada yang melakukan wisata ke Provinsi NTB karena adanya *event* yang diselenggarakan, sehingga wisatawan tertarik dengan *event* tahunan yang selalu diadakan oleh Pemerintah Provinsi NTB. Selain adanya *event* yang diadakan oleh pemerintah Prov. NTB wisatawan yang memilih datang ke Provinsi Nusa Tenggara Barat juga dikarenakan adanya momen libur sekolah baik itu libur panjang yang bertepatan dengan libur berderet dalam satu minggu ataupun libur panjang akhir semester dan libur pasca *Ramadhan*. Momen ini dijadikan waktu yang cocok untuk melakukan perjalanan wisata baik wisata perseorangan, kelompok ataupun wisata bersama keluarga.

Momen *event* tahunan dan libur sekolah ini bisa dijadikan *fitur* untuk diolah supaya bisa diketahui sejauh mana *fitur* ini bisa memberikan *performance* kinerja dalam melakukan prediksi wisatawan di Provinsi NTB. *Fitur* ini juga bisa dilakukan pencarian apakah *event* tahunan yang ada di NTB memiliki hubungan yang baik dengan libur sekolah sehingga bisa diketahui bahwa *event* yang ada sangat berhubungan dengan libur sekolah atau bahkan sebaliknya tidak memiliki hubungan dalam mendatangkan wisatawan ke Provinsi Nusa Tenggara Barat.

Dengan menggunakan metode *Long Short Term Memory (LSTM)* dan *Support Vector Regression (SVR)* penulis akan mencoba untuk melakukan prediksi terhadap jumlah wisatawan di daerah Provinsi Nusa Tenggara Barat. Dengan menggunakan metode *LSTM* dan *SVR* penulis akan mencoba melakukan pengujian terhadap variabel *independent* dan *dependent* sehingga menghasilkan berapa

besarkah nilai error dari variabel-variabel yang akan diuji terhadap prediksi yang dihasilkan, dan apakah prediksi yang didapatkan menggunakan variabel-variabel tersebut memiliki keterhubungan antar variabel atau tidak.

Menurut F.R.G.R. Siti Mujilawati (2017) Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi.

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, *LSTM* dikembangkan untuk melengkapi kekurangan dari *RNN* yang tidak dapat melakukan prediksi berdasarkan informasi masa lalu yang disimpan dalam jangka waktu yang lama. 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.

Prediksi menggunakan metode *Long Short Term Memory Network (LSTM)* sudah banyak digunakan dalam penelitian sebelumnya, penelitian yang dilakukan oleh Jian- Wu Bi, Yang Liu, dan Hui Li (2020) tentang *Daily tourism volume*

forecasting for tourist attractions. Pada penelitian tersebut menggunakan data dengan 3 atribut time series dalam melakukan peramalan yaitu, data *tourist volume* sebelumnya, data *search engine*, dan data cuaca, penelitian dilakukan di daerah wisata China yaitu Jiuzhaigou dan Area Pegunungan Huangshan. Adapun hasil prediksi diperoleh nilai *MAPE* terbaik dengan nilai 6,667% hasil ini menunjukkan bahwa metode *LSTM* dengan menggunakan 3 data time series sangat berpengaruh kuat dalam melakukan prediksi volume wisatawan dibandingkan tanpa hanya menggunakan satu data time series saja.

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.

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.

Sedangkan untuk metode *SVR* Ada beberapa peneliti yang telah melakukan prediksi wisatawan dengan menggunakan metode *support vector regression (SVR)*, seperti penelitian yang telah dilakukan oleh (Abellana Dkk, 2021) dengan

penelitian yang berjudul *Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology*, pada penelitian tersebut menghasilkan kesimpulan bahwa Model *hybrid SVR-SARIMA* yang diusulkan adalah model dengan kinerja terbaik di antara serangkaian model yang dipertimbangkan menggunakan kriteria kinerja yang mengevaluasi kesalahan besaran (*error*). Dengan hasil kinerja *SVR-SARIMA* kinerja *RMSE* menghasilkan 33749.73, untuk metode *Sarima* menghasilkan *RMSE* 142093.82, untuk metode *SVR* menghasilkan *RMSE* 61420.07 dan untuk hasil metode *HoltWinter* menghasilkan *RMSE* 105202.55.

Selain itu penelitian yang dilakukan oleh (Fan dkk, 2022) dengan penelitian yang berjudul *Application of COEMD-SSVR model in tourism demand forecasting and economic behavior analysis: The case of Sanya City* dalam penelitian tersebut Hasil menunjukan model yang diajukan (*COEMD- S-SVR*) lebih akurat dibandingkan dengan model lainnya, dengan hasil *MAPE* sebesar 0.0875%.

Selanjutnya penelitian yang dilakukan oleh (C. Li dkk, 2020) dengan penelitian yang berjudul *Forecasting tourist arrivals using denoising and potential factors*. Pada penelitian tersebut hasil menunjukkan bahwa denoising sangat meningkatkan akurasi peramalan, dan teknik denoising yang diusulkan mengungguli pendekatan lain. Selain itu, model yang diusulkan menunjukkan kinerja peramalan yang paling memuaskan diantara semua model benchmark, serta skalabilitas dan stabilitas yang sangat baik. Hasil *MAPE* dari beberapa metode sebagai berikut pengujian metode *ARIMA* menghasilkan *MAPE* sebesar 34.77%, pengujian dengan metode *ANN* menghasilkan kinerja *MAPE* sebesar 16.19%,

pengujian menggunakan metode *SVR* menghasilkan kinerja *MAPE* sebesar 12.11% dan hasil pengujian menggunakan metode gabungan *RHHT-SVR-PF* menghasilkan kinerja *MAPE* sebesar 7.45%.

Penelitian yang dilakukan oleh Listiya Surtiningsih¹, M. Tanzil Furqon, Sigit Adinugroho, 2018 tentang Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali Menggunakan *Support Vector Regression* dengan Algoritma Genetika. Pada penelitian tersebut hasil pengujian dengan nilai *MAPE* sebesar 2.513% dan parameter terbaik menghasilkan range *lamda* 1-10, range kompleksitas 1-100, range *epsilon* yaitu 0.00001-0.001 kemudian range *gamma* 0.00001-0.001, range *sigma* 0.01-3.5 jumlah iterasi *SVR* yaitu 1250, generasi GA 90 jumlah populasi 70 *crossover* rate 0.6, *mutation* rate 0.4 jumlah fitur 2 dan terakhir jumlah periode dalam melakukan prediksi adalah 1 bulan.

Penelitian yang dilakukan oleh L. Yao, R. Ma, and H. Wang, *Alexandria Engineering Journal*, 2021 dengan penelitian berjudul *Baidu index-based forecast of daily tourist arrivals through rescaled range analysis, support vector regression, and autoregressive integrated moving Average*. Menghasilkan kesimpulan bahwa diantara metode *R/S (Rescaled Range Analysis)*, *SVR*, dan *ARIMA*, metode *SVR* lebih unggul dalam prediksi nonlinear dan metode *ARIMA* lebih unggul dalam prediksi Linear. Dengan hasil masing-masing metode *RS-SVR-ARIMA* hasil *MAPE* 7.56% kemudian metode *RS-SVR* hasil *MAPE* 10.008% dan metode *SVR-ARIMA* hasil *MAPE* 10.218%

Proses dalam melakukan prediksi dengan menggunakan metode *Long Short Term Memory Network (LSTM)* menggunakan bahasa pemrograman *python* dengan

tools yang digunakan adalah *google colaboratory* atau *google colab*, *Google colaboratory* adalah suatu platform yang dikembangkan oleh *google research*, platform ini sangat cocok untuk keperluan machine learning dengan menggunakan kode *python*. Platform ini bersifat online yang artinya akses terhadap platform ini harus terhubung ke jaringan internet. Adapun untuk melakukan akses terhadap platform ini harus menggunakan browser dalam pengaksesannya Gabriel Irfon Elrohi Soen, Marlina, Renny (2022).

Nilai akurasi tingkat kesalan dalam peramalan digunakan untuk mengetahui sebesar mana tingkat kesalahan dalam melakukan peramalan M. Azman Maricar (2019). Terdapat banyak metode yang ada akan tetapi tidak semua metode bisa digunakan sesuai dengan kasus yang ada. Dalam melihat seberapa besar tingkat kesalahan ada tiga jenis metode secara umum *Mean Average Error (MAE)*, *Mean Absolute Deviation (MAD)*, *Mean Square Error (MSE)* dan *Mean Absolute Percent Error (MAPE)*

Dalam melakukan prediksi jumlah wisatawan dan melihat tingkat kesalahan (*error*) dibutuhkan suatu kumpulan data atau sering disebut dengan dataset atau juga basis data (*database*). Menurut Yahya dan Mahfuz (2019) Dataset adalah suatu basis data yang ada didalam memori. Dataset seperti halnya database memiliki semua karakteristik, fitur dan fungsi. Dataset memiliki banyak *record* yang berbeda beda sesuai dengan kebutuhan.

Penelitian ini bertujuan untuk melakukan prediksi terhadap wisatawan yang datang ke Provinsi Nusa Tenggara Barat, adapun dataset yang digunakan adalah dataset multivariate time series yang terdiri dari 3 atribut yaitu data volume

wisatawan tahun sebelumnya yang didapat dari webiste NTB Satu Data <https://data.ntbprov.go.id/search/type/dataset> dari tahun 2014 sampai dengan tahun 2022. Kemudian data *event* kegiatan di Provinsi NTB yang didapat dari Dinas Pariwisata Prov. Nusa Tenggara Barat dan juga data kalender liburan setiap tahun yang diambil dari tahun ajaran 2014/2015 dari *open data kalenderpendidikan.com*.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan diatas maka yang menjadi rumusan masalah utama dalam penelitian ini adalah :

- a. Berapakah tingkat kinerja dari metode *Long Short term Memory (LSTM)* dalam melakukan prediksi terhadap jumlah wisatawan dengan menggunakan variabel *event* dan variabel libur sekolah?
- b. Berapakah tingkat kinerja dari metode *Support Vector Regression (SVR)* dalam melakukan prediksi terhadap jumlah wisatawan menggunakan variabel *event* dan libur sekolah ?
- c. Metode manakah yang menghasilkan *performance* (kinerja) yang lebih baik dalam melakukan prediksi wisatawan menggunakan variabel libur sekolah dan variabel *event* ?

1.3. Batasan Masalah

Penulis melakukan pembatasan masalah agar penelitian lebih terfokus dan terarah, serta tidak meluas keluar dari penelitian sehingga tujuan penelitian akan tercapai. Adapun batasan masalah yang dimaksud dalam penelitian ini meliputi :

- a. Penelitian ini dilakukan dengan menggunakan data dari dataset publik yaitu data jumlah wisatawan Prov. NTB tahun 2014-2021, kemudian data agenda kegiatan bulanan dari Dinas Pariwisata Provinsi NTB tahun 2017-2022 dan data kalender pendidikan tahun ajaran 2014-2022.
- b. Algoritma yang akan digunakan dalam penelitian ini adalah *Long Short Term Memory Network (LSTM)* dan *Vector Support Regression (SVR)*.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang ada diatas, mak tujuan peneliti melakukan penelitian ini adalah sebagai berikut :

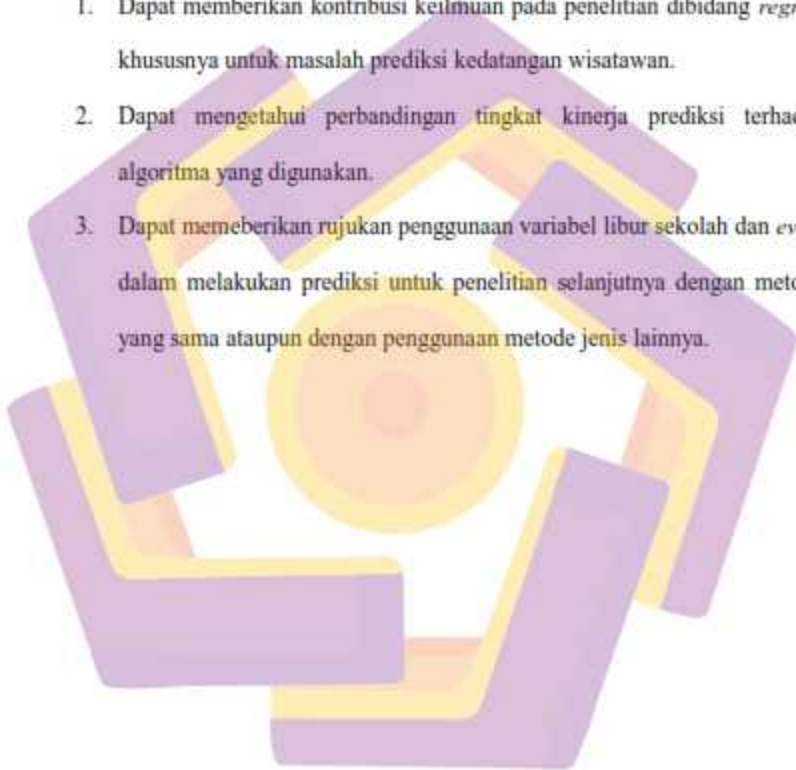
- a. Melakukan pengujian terhadap variabel libur sekolah dan *event* dengan menggunakan metode *Long Short Term Memory (LSTM)*.
- b. Melakukan pengujian terhadap variabel libur sekolah dan *event* dengan menggunakan metode *Support Vector Regression (SVR)*.

1.5. Manfaat Penelitian

Adapun manfaat yang diperoleh melalui penelitian yang dilakukan adalah sebagai berikut :

- a. Bagi Dinas Pariwisata Provinsi Nusa Tenggara Barat memberikan informasi terkait prediksi wisatawan dengan variabel yang telah digunakan, sehingga dapat membantu dalam persiapan menyambut wisatawan berdasarkan trend libur sekolah dan juga *event* tahunan.

- b. Bagi masyarakat dan ilmu pengetahuan, dapat memberikan pertimbangan tentang variabel yang digunakan sehingga variabel ini dapat digali lebih dalam untuk kebutuhan penelitian terkait prediksi.
- c. Manfaat Bagi peneliti yang didapat dari penelitian ini adalah sebagai berikut :
 - 1. Dapat memberikan kontribusi keilmuan pada penelitian dibidang *regresi* khususnya untuk masalah prediksi kedatangan wisatawan.
 - 2. Dapat mengetahui perbandingan tingkat kinerja prediksi terhadap algoritma yang digunakan.
 - 3. Dapat memeberikan rujukan penggunaan variabel libur sekolah dan *event* dalam melakukan prediksi untuk penelitian selanjutnya dengan metode yang sama ataupun dengan penggunaan metode jenis lainnya.



BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Dalam melakukan penelitian ini penulis terlebih dahulu melakukan tinjauan pustaka terhadap penelitian sebelumnya. Dari penelitian terdahulu penulis belum menemukan penggunaan variabel yang sama seperti yang digunakan pada penelitian ini baik dengan menggunakan metode yang sama ataupun dengan menggunakan metode yang berbeda. Akan tetapi penulis mengambil beberapa penelitian terdahulu dalam penggunaan beberapa metode penelitian yang nantinya akan digunakan sebagai referensi dan juga memperkaya bahan kajian oleh penulis dalam melakukan prediksi wisatawan, adapun metode yang akan dijadikan referensi pada penelitian terdahulu adalah metode *Long Short Term Memory (LSTM)* dan juga metode *Support Vector Regressor (SVR)*.

Berikut adalah beberapa penelitian yang terkait dengan masalah yang sedang diteliti :

A. Yifei Li dan Han Cao (2018)

Pada penelitian yang dilakukan oleh Yifei Li dan Han Cao dengan judul "*Prediction for Tourism Flow based on LSTM Neural Network*" (2018). Penelitian tersebut dibuat untuk melakukan prediksi terhadap kedatangan arus wisatawan dengan menggunakan metode *LSTM NN*. Optimasi yang digunakan dalam penelitian tersebut menggunakan *Adam Optimizer* dengan *learning rate adaptive*. Dataset yang digunakan untuk penelitian adalah dataset perjalanan

wisatawan yang diambil dari *small wild goose pagoda* di museum Xi'an. Dari Januari 2013 sampai dengan Desember 2015. Pada percobaan tersebut dataset dibagi menjadi dua yaitu 67% untuk data latih (*training data*) dan 33% digunakan untuk data uji (*test data*). Dengan hasil yang didapat bahwa metode LSTM NN menunjukkan hasil yang lebih baik dibandingkan dengan metode ARIMA dan BPNN. Dengan hasil kinerja menggunakan pengujian kinerja RMSE menghasilkan ARIMA 1488, BPNN 1365,35, Simplex LSTM 988,04, dan LSTM susun 1195,76.

B. Kajian He, Lei Ji, Chi, Wai Don Wu dan Kwok Fai Geoffrey Tso (2021)

Penelitian yang dilakukan oleh Kijian He dkk dengan judul "*Using SARIMA-CNN-LSTM approach to forecast daily tourism demand*" (2021). Pada penelitian tersebut bertujuan untuk melakukan prediksi terhadap data permintaan wisatawan pada frekuensi harian dengan menggunakan fitur linear dan non linear. Adapun dataset yang digunakan adalah data harian wisatawan di macao. Metode yang digunakan pada penelitian tersebut adalah metode SARIMA-CNN-LSTM, untuk menangkap fitur linier pada penelitian tersebut digunakan metode SARIMA, untuk menangkap struktur data hierarkies digunakan metode CNN, sedangkan untuk menangkap ketergantungan jangka panjang pada data digunakan metode LSTM. Hasil yang didapat dari pengeujian penelitian tersebut menghasilkan akurasi prediksi yang lebih besar dalam penggunaan metode SARIMA-CNN-LSTM dibandingkan menggunakan metode individual. Dengan hasil kinerja MAPE untuk pengujian SARIMA menghasilkan MAPE sebesar 0,0999%, unum CNN menghasilkan

MAPE sebesar 0,1355%, untuk LSTM menghasilkan MAPE sebesar 0,1195% dan untuk pengujian dengan menggunakan SARIMA-CNN-LSTM menghasilkan 0,0986% dengan dataset sample yang digunakan adalah kedatangan dari negara Cina ke Macau.

C. Ahmad Ashril Rizal dan Siti Soraya (2018)

Penelitian yang dilakukan oleh Ahmad Ashril Rizal dan Siti Soraya dengan judul "*Multi Time Steps Prediction Dengan Recurrent Neural Network Long Short Term Memory*" (2018). Pada penelitian tersebut bertujuan untuk melakukan prediksi terhadap kunjungan wisatawan dengan menggunakan metode *Recurrent Neural Network Long Short Term Memory (RNN LSTM)*. Dataset yang digunakan adalah kunjungan wisatawan berdasarkan history sebelumnya dari tahun 2010 sampai dengan tahun 2016. Dataset yang ada dilakukan pembagian data *training* dan data *testing* dengan pembagian 67% data *training* dan 33% data *testing*. Hasil yang didapat dari pengujian data *training* menghasilkan nilai dengan pengujian kinerja RMSE sebesar 6529,42 sementara hasil pengujian data *testing* menghasilkan nilai RMSE sebesar 13766,85.

D. Harun Mukhtar, Yoze Rizki, Febby Apri Wenando, Muhammad Abdul Al Aziz (2022)

Penelitian yang dilakukan oleh Harun Mukhtar dkk dengan judul "*Prediksi Kunjungan Wisatawan dengan Reduksi Noise pada Google Trends menggunakan Hilbert-Huang Transform dan Long Short-Term Memory*" (2022). Pada penelitian tersebut dataset yang digunakan adalah data pencarian

dari *google trends* dengan 7 *query* yaitu : Indonesia, *tour to Indonesia*, Bali, Query rekomendasi Bali Indonesia, hotel Bali Indonesia dan Best hotel in Bali Indonesia. Data yang digunakan adalah data dari 01-01-2008 sampai dengan 31-12-2021. Dataset selanjutnya adalah dataset yang diambil dari data Badan Pusat Statistik Indonesia yang berisi data jumlah kunjungan wisatawan perbulan menurut pintu masuk dari mulai 01-01-2008 sampai dengan 31-12-2021. Kemudian dataset yang ada dilakukan pemrosesan Noise dengan menggunakan *Hilbert-Huang Transform (HHT)*. Selanjutnya data yang ada dilakukan pengujian, pengujian menggunakan data sebanyak 156 baris dari tahun 2009 sampai dengan tahun 2019 menghasilkan nilai RMSE sebesar 129249 dengan menggunakan LSTM. Sedangkan pengujian menggunakan HHT+LSTM menghasilkan RMSE sebesar 653058.

E. Dharyll Prince Mariscal Abellana, Donna Marie Canizares Rivero, Ma. Elena Aparente dan Arles Rivero Abellana (2021)

Penelitian yang dilakukan oleh Dharyll Prince Mariscal Abellana dkk dengan penelitian berjudul "*Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: a Philippine scenario*" (2021). Pada penelitian tersebut bertujuan untuk melakukan peramalan permintaan pariwisata jangka panjang di daerah Filipina. Metode yang digunakan dalam melakukan peramalan adalah metode hybrid yang menggabungkan antara metode SVR dan metode SARIMA, dalam penelitian ini juga merupakan penelitian yang pertama kali mengusulkan dan mendemonstrasikan penggunaan pendekatan *Multiple Criteria Decision-*

Making (MCDM). Dataset yang digunakan dalam penelitian adalah dataset histori kunjungan wisatawan dari tahun 1998 sampai dengan tahun 2018. Dataset tersebut kemudian dilakukan *splitting* menjadi data *training* dan data *testing*, selanjutnya pengujian dilakukan pengukuran tingkat kinerja model dengan menggunakan MAPE, adapun hasil dari penelitian tersebut menghasilkan nilai MAPE dengan pengujian menggunakan metode SVR sebesar 10,15%, kemudian pengujian dengan menggunakan metode SARIMA menghasilkan nilai MAPE sebesar 27,88% sedangkan pengujian menggunakan metode *hybrid* dengan menggabungkan metode SVR-SARIMA menghasilkan nilai MAPE sebesar 5,45%.

F. Guo-Feng Fan, Xiang-Ru Jin dan Wei-Chlang Hong (2021)

Penelitian yang dilakukan oleh Guo-Feng Fan dkk dengan penelitian berjudul "*Application of COEMD-S-SVR model in tourism demand forecasting and economic behavior analysis: The case of Sanya City*" (2021). Pada penelitian ini bertujuan untuk melakukan peramalan terhadap permintaan wisatawan di Tiongkok khususnya di kota Sanya, dataset yang digunakan adalah dataset histori wisatawan di kota Sanya China dari tahun 2012 sampai dengan tahun 2018 dengan jumlah data sebanyak 78 data. Hasil dari pengujian dengan menggunakan metode ARIMA menghasilkan MAPE sebesar 0,0352%, pengujian menggunakan metode BPNN menghasilkan MAPE sebesar 0,1939% pengujian menggunakan metode SVR menghasilkan MAPE sebesar 0,0334%, sedangkan pengujian menggunakan metode SVR-COEMD-S menghasilkan nilai MAPE sebesar 0,1853%.

G. Listya Surtiningsih, M. Tanzil Furqon dan Sigit Adinugroho (2018)

Penelitian yang dilakukan oleh Listya Surtiningsih dkk dengan penelitian berjudul “Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali Menggunakan *Support Vector Regression* dengan Algoritma Genetika” (2018). Pada penelitian tersebut bertujuan untuk melakukan prediksi terhadap kunjungan wisatawan mancanegara dengan menggunakan metode peralaman *Support Vector Regressor (SVR)* dengan menggunakan metode Optimasi Algoritma Genetika (GA) dalam mendapatkan parameter terbaik. Hasil dari penelitian tersebut menghasilkan nilai kinerja MAPE sebesar 2,513% dan parameter terbaik menghasilkan range lamda 1-10, range kompleksitas 1-100, range epsilon yaitu 0.00001-0.001 kemudian range gamma 0.00001-0.001, range sigma 0.01-3.5 jumlah iterasi SVR yaitu 1250, generasi GA 90 jumlah populasi 70 crossover rate 0.6, muttation rate 0.4 jumlah fitur 2 dan terakhir jumlah periode dalam melakukan prediksi adalah 1 bulan.

H. Rui Zhang, dkk (2022)

Penelitian yang dilakukan oleh Rui Zhang, dkk (2022) dengan judul. “*A Comparison of LSTM and Linear Regression for Non-Linear Time Series Prediction*” pada penelitian yang mereka lakukan menunjukan hasil pengujian menggunakan metode ARIMA tampil lebih baik dalam kejadian bulanan dan mingguan akan tetapi jika data yang digunakan menggunakan data harian yang mana data yang digunakan otomatis lebih banyak maka metode LSTM tampil lebih unggul dibanding dengan metode ARIMA. Dalam penelitian yang mereka lakukan juga menyatakan bahwa semakin banyak dataset yang

digunakan maka metode LSTM akan tampil lebih baik dibanding dengan metode lain yang mereka gunakan.

I. Ghahreman Abdoli, Mohsen Mehr Ara & Mohammad Ebrahim Ardalani, 2020

penelitian yang dilakukan oleh Ghahreman Abdoli dkk dengan judul "*Comparing The Prediction Accuracy Of Lstm And Arima Models For Time-Series With Permanent Fluctuation*" pada penelitian tersebut metode LSTM mengungguli metode lain yang digunakan pada penelitian tersebut. Dari berbagai pengujian yang dilakukan LSTM selalu mengungguli model ARIMA dengan hasil akurasi yang lebih baik.

J. Gourav Bathla, 2020

Penelitian yang dilakukan oleh Gourav Bathla dengan judul "*Stock Price prediction using LSTM and SVR*" eksperimen yang telah dilakukan menunjukan bahwa metode LSTM dalam melakukan prediksi lebih baik dibanding dengan metode SVR, akurasi yang dihasilkan oleh LSTM mengungguli metode SVR, penelitian yang dilakukan mengukur kinerja menggunakan MAPE sebagai pengukuran tingkat error prediksi, hasil yang didapat dalam memprediksi setiap saham yang diuji semua pengujian menghasilkan metode LSTM lebih unggul daripada SVR.

K. Ahmad Kamal Badri dkk, 2022

Pada Penelitian yang dilakukan oleh Ahmad Kamal Badri dkk dengan judul "*Decision-Making Techniques using LSTM on Antam Mining Shares before and during the COVID-19 Pandemic in Indonesia*". Pada penelitian

tersebut model yang digunakan dalam melakukan penelitian adalah LSTM dengan dataset dua kondisi yaitu kondisi ketika covid-19 dan juga saat kondisi normal. Hasil pengujian menunjukkan bahwa model LSTM dapat dengan baik digunakan dalam melakukan prediksi menggunakan dataset historis.

L. YOGI FIRGIAWAN, 2023

Pada penelitian yang dilakukan oleh Yogi Firgiawan dengan judul "Prediksi Jumlah Produksi Perkebunan Besar di Indonesia Menggunakan Algoritma Long Short Term Memory dan Support Vector Regression" pada penelitian tersebut bertujuan untuk melakukan prediksi terhadap minyak sawit, karet kering, dan teh menggunakan metode LSTM dan SVR. Hasil yang didapatkan dari penelitian yang telah dilakukan bahwa metode SVR tampil lebih baik dibanding dengan model LSTM.

2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian
Pengujian Variabel Event Tahunan Dan Libur Sekolah Dalam Memprediksi Jumlah Wisatawan (Studi Kasus: Provinsi Nusa Tenggara Barat)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Prediction for Tourism Flow based on LSTM Neural Network	Yifei Li dan Han Cao, <i>Procedia Computer Science</i> , 2018	Melakukan prediksi terhadap kedatangan arus wisatawan dengan menggunakan metode LSTM NN	Hasil pengujian dengan <i>data splitting</i> 67% <i>data training</i> dan 33% <i>data testing</i> menghasilkan nilai kinerja RMSE untuk masing-masing metode ARIMA sebesar 1488, BPNN sebesar 1365,35, <i>simplex LSTM</i> 988,04 dan <i>LSTM</i> susun 1195,76 hal ini menunjukan bahwa metode LSTM NN menunjukan hasil yang lebih baik, efektif dan lebih sederhana dibandingkan dengan metode lainnya.	Untuk penelitian selanjutnya dimasa depan perlu mempertimbangkan berbagai faktor untuk pariwisata seperti, suhu rata-rata, data cuaca, jumlah hari dalam seminggu dan faktor lainnya.	Perbedaan peneliti sebelumnya berada pada jenis variabel yang digunakan, pada penelitian sebelumnya variabel yang digunakan hanya menggunakan histori kedatangan wisatawan dan pada penelitian yang saya lakukan variabel yang digunakan adalah histori wisatawan, <i>event</i> tahunan daerah dan kalender libur sekolah. metode yang digunakan dalam penelitian sebelumnya menggunakan metode <i>hybrid SARIMA-CNN-LSTM</i> sedangkan pada penelitian yang saya lakukan menggunakan metode LSTM dan juga metode SVR.

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2	Using SARIMA--CNN-LSTM approach to forecast daily tourism demand	Kaijian He, Lei Ji, Chi, Wai Don Wu dan Kwok Fai Geoffrey Tso, Journal of Hospitality and Tourism Management, 2020	Melakukan prediksi terhadap data permintaan wisatawan frekuensi harian dengan menggunakan fitur <i>linear</i> dan <i>non linear</i>	Hasil pengujian kinerja MAPE dengan menggunakan metode SARIMA menghasilkan nilai MAPE sebesar 0,0999%, dengan menggunakan metode CNN menghasilkan nilai MAPE sebesar 0,1355% dan untuk LSTM menghasilkan MAPE sebesar 0,1195% sedangkan hasil pengujian menggunakan metode sebesar 0,1195% sedangkan hasil pengujian menggunakan metode <i>hybrid SARIMA-CNN-LSTM</i> menghasilkan nilai MAPE sebesar 0,0986%. Hal ini menunjukan bahwa kinerja dari metode <i>hybrid</i> lebih baik daripada metode individu.	Penelitian berdasarkan data deret waktu yang lebih panjang dan data dari negara tujuan lain adalah arah yang memungkinkan untuk penelitian di masa depan.	Variabel yang digunakan pada penelitian sebelumnya hanya menggunakan data time series histori wisatawan saja sedangkan pada penelitian yang saya lakukan variabel yang digunakan adalah histori wisatawan, <i>event</i> tahunan daerah, kalender libur sekolah, metode yang digunakan dalam penelitian sebelumnya menggunakan metode <i>hybrid SARIMA-CNN-LSTM</i> sedangkan pada penelitian yang saya lakukan menggunakan metode LSTM dan juga metode SVR

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	Multi Time Steps Prediction Dengan Recurrent Neural Network Long Short Term Memory	Ahmad Ashril Rizal dan Siti Soraya, Jurnal Matrix, 2018	Melakukan prediksi terhadap kunjungan wisatawan menggunakan metode RNN LSTM	Hasil yang didapat dari pengujian data training menghasilkan nilai RMSE sebesar 6529,42 dan pengujian data testing menghasilkan nilai RMSE sebesar 13766,85, pengujian menggunakan tiga model LSTM yang beragam hasil terbaik adalah dengan menggunakan model LSTM regresi dengan nilai RMSE 6529,42	Penelitian dapat dilakukan dengan menggunakan metode yang lain, diharapkan dengan menggunakan metode yang lain dapat memberikan nilai akurasi yang lebih baik, selain itu penggunaan optimasi penentuan bobot lain dapat menjadikan proses berjalan lebih cepat.	Variabel yang digunakan dalam penelitian sebelumnya menggunakan data time series histori wisatawan, sedangkan pada penelitian yang saya lakukan variabel yang digunakan menggunakan variabel histori wisatawan, <i>event</i> tahunan dan kalender libur sekolah, metode yang digunakan dalam penelitian sebelumnya menggunakan RNN LSTM sedangkan pada penelitian yang saya lakukan menggunakan metode LSTM dan SVR.
4	Prediksi Kunjungan Wisatawan dengan Reduksi Noise pada Google Trends menggunakan Hilbert-Huang Transform dan Long Short-Term Memory	Harun Mukhtar, Yoze Rizki, Febby Apri Wenando, Muhammad Abdul Al Aziz, Jurnal Fasilkom (2022)	Melakukan prediksi wisatawan menggunakan variabel pencarian <i>google trends</i> dengan pemrosesan <i>Noise</i> menggunakan Hilbert Huang Transform (HHT).	Dataset hasil HHT tidak dapat digunakan seluruhnya dikarenakan terjadi <i>error</i> yang sangat besar dan berlawanan dengan <i>trends</i> pada Januari tahun 2020 sampai desember tahun 2021, maka data pada periode	Disarankan untuk menambah <i>query</i> dari <i>google trends</i> sehingga hasil perhitungan korelasi lebih baik. Disarankan untuk menggunakan dataset dengan rentang waktu perhari.	Pada penelitian sebelumnya menggunakan variabel <i>google trends</i> sedangkan pada penelitian yang saya lakukan menggunakan variabel <i>event</i> dan libur sekolah.

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				tersebut tidak digunakan atau dibuang. Pengujian dilakukan menggunakan dataset sebanyak 156 data dan menghasilkan nilai RMSE LSTM sebesar 129249 sedangkan hasil dari HHT-LSTM menghasilkan RMSE sebesar 653058		
5	Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: a Philippine scenario	Dharyll Prince Mariscal Abellana, Donna Marie Canizares Rivero, Ma. Elena Aparente dan Arics Rivero Abellana (2021)	Mengusulkan model peramalan gabungan (<i>hybrid</i>) untuk peramalan permintaan jangka panjang di Filipina menggunakan metode SVR-SARIMA	Temuan model <i>hybrid</i> SVR-SARIMA adalah model kinerja terbaik dibandingkan dengan model yang lainnya. MAPE menghasilkan 10,15% untuk metode SVR, 27,88% untuk metode SARIMA dan 5,445% untuk metode <i>hybrid SVR-SARIMA</i> . Pada penelitian ini juga mengusulkan pendekatan MCDM untuk pemilihan model terbaik, dan metode	Penelitian menggunakan metode MCDM dilakukan tanpa membandingkan dengan pendekatan MCDM lainnya, dan juga pendekatan optimasi <i>hyperparameter</i> belum dilakukan untuk menyesuaikan model <i>hybrid</i> yang diusulkan.	Pada penelitian tersebut metode yang digunakan menggunakan metode SVR-SARIMA sedangkan pada penelitian yang saya lakukan menggunakan metode SVR dan LSTM, variabel yang digunakan dalam penelitian sebelumnya hanya menggunakan histori kedatangan wisatawan diwaktu sebelumnya, sedangkan yang saya gunakan adalah variabel, histori data, event dan juga liburan sekolah.

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				PROMETHEE II digunakan untuk mengurutkan model permalan menggunakan fungsi preferensi.		
6	Application of COEMD-S-SVR model in tourism demand forecasting and economic behavior analysis: The case of Sanya City	Guo-Feng Fan, Xiang-Rui Jin dan Wei-Chiang Hong (2021)	Melakukan peramalan terhadap permintaan wisatawan di Tiongkok khususnya kota Sanya menggunakan dataset histori wisatawan dari tahun 2012 sampai dengan tahun 2018. Metode yang digunakan adalah SVR-COEMD-S	Hasil dari pengujian dengan menggunakan metode ARIMA menghasilkan MAPE sebesar 0,0352%, pengujian menggunakan metode BPNN menghasilkan MAPE sebesar 0,1939% pengujian menggunakan metode SVR menghasilkan MAPE sebesar 0,0334%, sedangkan pengujian menggunakan metode SVR-COEMD-S menghasilkan nilai MAPE sebesar 0,1853%	Disarankan agar studi kedepannya lebih komprehensif disarankan untuk melakukan penelitian pada faktor lain dan mempelajari pengaruhnya terhadap volatilitas kedatangan wisatawan bulanan.	Penggunaan variabel pada penelitian sebelumnya menggunakan variabel histori wisatawan saja, sedangkan pada penelitian yang saya lakukan menggunakan variabel histori, event dan juga liburan sekolah.

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
7	Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali Menggunakan Support Vector Regression dengan Algoritma Genetika	Listiya Surtiningsih, M. Tanzil Furqon dan Sigit Adinugroho (2018)	Melakukan prediksi terhadap kunjungan wisatawan mancanegara dengan menggunakan metode peraluman Support Vector Regressor (SVR) dengan menggunakan metode Optimasi Algoritma Genetika (GA) dalam mendapatkan parameter terbaik	Hasil dari penelitian tersebut menghasilkan nilai kinerja MAPE sebesar 2,513% dan parameter terbaik menghasilkan range lamda 1-10, range kompleksitas 1-100, range epsilon yaitu 0.00001-0.001 kemudian range gamma 0.00001-0.001, range sigma 0.01-3.5 jumlah iterasi SVR yaitu 1250, generasi GA 90 jumlah populasi 70 crossover rate 0.6, mutation rate 0.4 jumlah fitur 2 dan terakhir jumlah periode dalam melakukan prediksi adalah 1 bulan.	Saran yang dapat dijadikan pertimbangan adalah mengembangkan penelitian yang telah dilakukan dalam hal penggunaan metode optimasi Algoritma genetika bisa diganti dengan metode optimasi yang lain seperti PSO, ACO dan lainnya serta penggunaan kernel yang berbeda selain RBF.	Penelitian sebelumnya menggunakan variabel histori wisatawan saja, sedangkan penelitian yang saya lakukan menggunakan variabel histori wisatawan, event dan juga libur sekolah.
18	A Comparison of LSTM and Linear Regression for Non-Linear Time Series Prediction	Rui Zhang, Hejia Song, Qiulan Chen, Yu Wang, Songwang Wang dan Yonghong Li.	Melakukan prediksi kejadian demam berdarah di Tiongkok berdasarkan kejadian bulanan, mingguan dan	dari hasil penelitian yang dilakukan model ARIMA terpilih sebagai model yang paling sesuai untuk memprediksi kejadian	Dalam penelitian menggunakan metode LSTM dataset yang digunakan hanya sekitar 84 sampel untuk data bulanan, hal ini yang	Tujuan penelitian pada penelitian ini untuk memprediksi kejadian demam berdarah sedangkan tujuan yang saya lakukan melakukan pengujian

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
		PLoS ONE, 2022	harian, menggunakan metode ARIMA dan LSTM menggunakan dataset dari tahun 2013 sampai dengan 2018.	bulanan dan mingguan sedangkan LSTM tampil lebih baik pada kejadian harian.	menyebabkan model ARIMA tampil lebih baik sehingga disarankan untuk penelitian dengan model LSTM menggunakan dataset yang lebih banyak lagi.	variabel dalam melakukan prediksi wisatawan, metode yang digunakan sebagai pembandingan pada penelitian ini adalah metode ARIMA sedangkan metode yang saya gunakan sebagai pembandingan pada penelitian ini adalah SVR.

2.3. Landasan Teori

2.3.1. Prediksi/Peramalan

Peramalan merupakan salah satu alat bantu yang cukup efektif dan efisien dalam membuat suatu perencanaan dimasa yang akan datang (Nafis Sultan dkk, 2020). Ada dua jenis peramalan yaitu peramalan subjektif dan juga peramalan objektif, model yang digunakan dalam peramalan subjektif adalah model kualitatif, model ini akan sangat berguna dan bermanfaat ketika data kuantitatif tidak akurat atau data kuantitatifnya sulit untuk didapatkan. sedangkan peramalan objektif menggunakan model kuantitatif, metode objektif ini juga terbagi menjadi dua model yaitu model kausa dan model Time Series (Listya Surtiningsih dkk, 2018). Sudah bertahun-tahun berbagai model deret waktu (*time series*) telah banyak dilakukan untuk meramalkan kedatangan wisatawan (Holt, 1957; Hasani Dkk, 2015)

2.3.2 Variabel

Variabel penelitian menurut (Silaen, 2018) adalah suatu konsep yang memiliki bermacam-macam nilai yang berbeda/bervariasi, yaitu suatu sifat, karakteristik atau fenomena yang bisa menunjukkan suatu hal yang dapat diamati dan diukur dengan hasil nilai yang berbeda-beda. Sedangkan menurut (Sugiyono, 2019:68) variabel adalah suatu sifat atau atribut berupa nilai dari seseorang atau objek bahkan suatu kegiatan yang memiliki nilai dan mempunyai variasi tertentu yang ditetapkan oleh seorang peneliti untuk dapat

dipelajari dan kemudian dapat ditarik kesimpulannya. Pada penelitian yang penulis lakukan variabel yang digunakan terdiri dari variabel *independent* yaitu variabel *event* dan libur sekolah dan juga variabel *dependent* yaitu variabel histori wisatawan.

2.3.3. Preprocessing MinMax-Scaler

Dataset yang telah dimiliki harus dilakukan *preprocessing* agar nilai variabel dalam dataset menjadi rentang yang lebih terstandarisasi (K. Dewi Dkk, 2018). Dalam melakukan *preprocessing* data ada dua tahapan yang harus dilakukan yaitu tahapan pembersihan data dan tahapan normalisasi data (Rahmad Firdaus dan H. Mukhtar, 2023). Dalam melakukan normalisasi dataset pada penelitian ini menggunakan normalisasi *MinMaxScaler*, teknik ini merubah nilai yang ada menjadi nilai *interval* antara 0 dan 1 (D. Heksaputra, 2020). Adapun persamaan nya seperti pada persamaan 1 berikut.

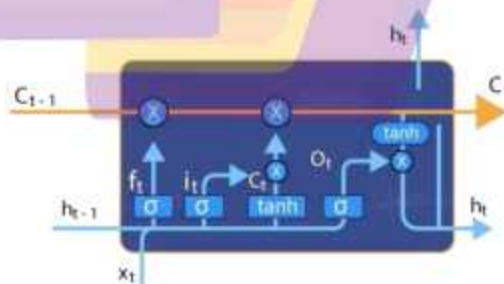
$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

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} merupakan nilai maksimum dari setiap kolom dalam *matrix dataset* (H. Mukhtar dkk, 2022).

2.3.4. Long Short Term Memory (LSTM)

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 (Hidayat R. F. Dkk, 2023), *LSTM* dikembangkan untuk melengkapi kekurangan dari *RNN* yang tidak dapat melakukan prediksi berdasarkan informasi masa lalu yang disimpan dalam jangka waktu yang lama (J. Manulang Dkk, 2020). 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 (A. Nilsen, 2022). Perbedaan yang paling utama antara *RNN* dan *LSTM* adalah pada *LSTM* memiliki empat lapisan modul jaringan dan informasi dipertukarkan dengan cara khusus. Struktur *LSTM* sendiri memiliki beberapa gerbang dengan fungsi dan tugasnya masing-masing, yaitu *Forget Gate* bertugas untuk melupakan beberapa informasi yang sudah tidak relevan dan sudah tidak dibutuhkan oleh sebuah sistem. *Input gate* bertugas untuk memasukan informasi yang berguna untuk mendukung keakuratan data. *Output gate* berfungsi untuk menghasilkan data yang komplit dan aktual.



Gambar 2.1 Kondisi Sel Modul LSTM

Sumber: intelliguat.com

Jaringan LSTM dirancang untuk melakukan penyimpanan informasi masa lalu jangka panjang dengan hasil kinerja yang sangat baik. Inti jaringan dari LSTM dapat dilihat di gambar 2. Pada gambar 2 terdapat garis *horizontal* antara C_{t-1} dan C_t informasi dapat dengan mudah mengalir disepanjang garis *horizontal* tersebut.

Prediksi jumlah wisatawan dengan menggunakan data histori wisatawan telah terbukti memiliki tren yang tinggi, dan variabel histori wisatawan telah menjadi prediktor umum yang digunakan dalam melakukan prediksi wisatawan. Namun, beberapa penelitian menyarankan penggunaan variabel prediktor tambahan guna meningkatkan akurasi prediksi (Y. P. Adiwiguna dkk, 2022). Salah satu variabel tambahan yang dapat digunakan adalah data *event* yang diadakan di suatu daerah secara bertahap, karena menjadi sumber informasi pariwisata yang signifikan (S. Lionetti dkk, 2021). Selain itu, data libur juga memiliki dampak yang besar terhadap jumlah wisatawan, terutama dengan adanya kegiatan mengisi waktu libur yang ramai dalam setiap tahunnya (N. A. Jauhariyah, dkk, 2021). Pada penelitian sebelumnya, metode *linear* telah digunakan untuk melakukan prediksi menggunakan variabel kalender libur, pada penelitian ini metode yang digunakan menggunakan metode terkini yaitu *Long Short Term Memory (LSTM)*.

2.3.5. Support Vector Regression (SVR)

Metode SVR merupakan metode penerapan dari metode *Support*

Vector Machine (SVM) untuk kasus penanganan masalah regresi. Metode ini dapat mengatasi *overfitting* karena bisa mengestimasi suatu fungsi dengan meminimalisir batas atas dari keadaan error secara umum atau sering disebut dengan konsep *risk minimization*. Hasil output dari penerapan metode SVR akan menghasilkan bilangan dalam bentuk riil (Rien Difitria, dkk, 2020). Adapun langkah-langkah dalam melakukan prediksi pertama melakukan inisialisasi terhadap parameter yang akan digunakan diantaranya parameter kompleksitas (C), iterasi maksimal, parameter *epsilon* (ϵ), *cLR*, *lamda* (λ), parameter *sigma* (α), inisialisasi *lag range*. Langkah selanjutnya adalah menghitung nilai *kernel* yang digunakan lalu kemudian menghitung nilai kesalahan (*error*) menggunakan pengukuran kinerja yang digunakan (Rien Difitria, Imam Cholissodin, Indriati, 2020).

SVR menggunakan pendekatan *geometris* dan teori tentang vektor dukungan untuk memodelkan hubungan antara variabel *input* (atribut) dan variabel *output* (target) dalam data (J. Chung dkk, 2014). SVR mencoba untuk menemukan *hyperplane* (bidang pembatas) yang memiliki *margin* maksimum di antara titik data dan memenuhi batasan toleransi kesalahan. Tujuan utamanya adalah untuk meminimalkan kesalahan regresi dengan mempertimbangkan kesalahan yang masih berada di dalam batas toleransi. SVR mampu menangani data yang tidak linear melalui transformasi nonlinear menggunakan fungsi kernel. Dalam metode SVR fungsi kernel memiliki dua jenis yang sering digunakan yaitu fungsi *kernel linear* dan fungsi *kernel polynomial*. Dalam *machine learning* sendiri biasanya jika teknik dalam

machine learning yang sedang diteliti menggunakan asumsi kelinieran maka metode atau algoritma yang digunakan juga terbatas untuk kasus yang *linear* (D. Saputra dan M. Ari, 2020). Adapun persamaan dari metode *Support Vector Regression (SVR)* dapat dilihat sebagai berikut ini :

Persamaan SVR Linear

$$Y = \sum_{i=1}^N (a_i - a^*_i) \cdot (x_i, x) + b \quad (2)$$

Pada persamaan diatas i merupakan *indeks training* dari 1 hingga N , a_i koefisien *lagrange* yang diperoleh selama proses *optimasi SVR*, a^*_i merupakan koefisien *lagrange* ganda yang diperoleh selama proses *optimasi SVR*. x_i, x perkalian titik antara vektor fitur *data training* x_i dan vektor fitur *data input* x . Dan b adalah bias atau konstanta.

Persamaan SVR Non Linear

$$Y = \sum_{i=1}^N (a_i - a^*_i) \cdot K(x_i, x) + b \quad (3)$$

Pada persamaan diatas i merupakan *indeks training* dari 1 hingga N , a_i koefisien *lagrange* yang diperoleh selama proses *optimasi SVR*, a^*_i merupakan koefisien *lagrange* ganda yang diperoleh selama proses *optimasi SVR*. $K(x_i, x)$ *kernel function* yang menghitung kesamaan atau jarak antara *data training* x_i dan *data input* x , dan b adalah bias konstanta.

2.3.6. Pengukuran Kinerja

Kinerja model peramalan sangat penting untuk dievaluasi sebagai tahap validasi model yang cocok untuk data penelitian (Maftahatul Hakimah Dkk, 2020). Kinerja model peramalan dilihat berdasarkan kesalahan peramalan yang dihasilkan. Kesalahan model peramalan ada berbagai jenis diantaranya :

1. Mean Absolute error (MAE)

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (3)$$

2. Mean Absolute Deviation (MAD)

$$MAD = \frac{1}{n} \sum_{t=1}^n |et| \quad (4)$$

dengan et : selisih antara nilai aktual dengan nilai ramalan pada saat periode t

3. Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|et|}{y} \times 100 \quad (5)$$

4. Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (et)^2} \quad (6)$$

2.3.7. Dataset

2.3.7.1. Dataset publik NTB Satu Data

NTB Satu Data merupakan data terbuka yang menyediakan data resmi yang bersangkutan dengan pembangunan Provinsi Nusa Tenggara Barat yang dihimpun dari berbagai organisasi perangkat daerah dan

pemangku kepentingan lainnya. Data ini tersedia dalam bentuk *excel* (.xlsx) yang bisa langsung didownload pada laman resmi NTB Satu Data yaitu pada link <https://data.ntbprov.go.id/>. Pada file data ini terdiri dari kolom bulan dalam satu tahun kemudian kolom jumlah wisatawan lokal dan kolom jumlah wisatawan mancanegara. Setiap baris berisi riwayat data pengunjung atau wisatawan yang berbeda untuk setiap bulannya. Data yang tersedia dalam setiap tahun adalah 12 bulan sehingga data yang ada untuk tahun 2014 sampai dengan 2022 adalah $9 \text{ tahun} \times 12 \text{ bulan} = 108$ Data wisatawan lokal dan 108 data wisatawan mancanegara. Untuk lebih jelas mengenai dataset NTB Satu data dapat dilihat pada lampiran dataset.

2.3.7.2 Data Calender Of Event NTB

Data ini digunakan untuk dijadikan variabel dalam melakukan prediksi wisatawan berdasarkan *event* yang terjadi pada tiap bulan pertahun yang ada di Prov Nusa Tenggara Barat. Adapun data didapat dari situs resmi dinas pariwisata Prov. Nusa Tenggara Barat yaitu pada halaman berikut <http://www.dishbudpar.ntbprov.go.id/promosi-ntb/>

2.3.7.3 Data Liburan Sekolah

Data ini digunakan untuk dijadikan variabel dalam melakukan prediksi wisatawan berdasarkan jumlah libur sekolah setiap bulannya. Data diambil pertahun dari tahun 2014 sampai dengan tahun 2022 yang ada di Prov. Nusa Tenggara Barat. Adapun data didapat dari situs publik yaitu pada halaman berikut <https://kalenderpendidikan.com>.

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Pada penelitian ini menggunakan jenis penelitian *Eksperimen*, dimana peneliti melakukan investigasi sebab akibat dengan menggunakan ujicoba yang dikontrol oleh peneliti. Pada penelitian ini penerapan variabel baru akan digunakan dalam pengujian yang akan diterapkan menggunakan metode LSTM dan SVR. Dalam hal ini peneliti akan melakukan *eksperimen* terhadap penggunaan variabel *event* dan juga variabel libur sekolah terhadap pengaruh prediksi menggunakan metode LSTM dan SVR.

Dalam melakukan penelitian sifat penelitian yang digunakan adalah penelitian deskriptif, dimana peneliti berusaha menjawab permasalahan yang dengan data-data dengan proses penyajian data, melakukan analisis data dan juga menginterpretasikannya (Narbuka & Ahmadi :2015).

Adapun pendekatan yang digunakan pada penelitian ini adalah pendekatan kuantitatif dimana penelitian yang dilakukan dari mulai pengumpulan data, proses analisis sampai dengan hasilnya mempergunakan aspek perhitungan dan kepastian data numerik.

3.2. Metode Pengumpulan Data

Dalam melakukan penelitian ini penulis melakukan pengumpulan data dengan menggunakan jenis pengumpulan data sekunder. Pengumpulan data sekunder dilakukan secara tidak langsung kepada pihak atau objek yang diteliti.

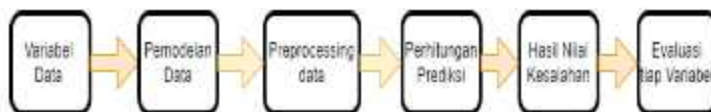
Peneliti mendapatkan data yang sudah jadi yang dikumpulkan oleh pihak lain dengan berbagai metode baik metode komersial ataupun non komersial baik yang bisa diakses secara daring ataupun secara luring. Contoh pengumpulan data jenis sekunder adalah dengan melakukan penelitian melalui penelitian kepustakaan dimana peneliti mengumpulkan data dengan mempelajari buku, jurnal, literatur dan artikel lain yang berkaitan dengan masalah yang sedang diteliti. Pada kasus ini penulis melakukan pengumpulan data dari berbagai sumber literatur sehingga dalam melakukan penelitian ini sistem pengumpulan data yang dilakukan adalah jenis pengumpulan data sekunder.

Adapun untuk pengumpulan data penelitian penulis melakukan beberapa langkah sebagai berikut :

1. Pengambilan dataset time series wisatawan pada laman resmi NTB satu data secara online pada link <https://data.ntbprov.go.id/dataset/jumlah-kunjungan-wisatawan-ke-provinsi-nusa-tenggara-barat-ntb>
2. Pengambilan dataset Data Calendar of Event kalender kegiatan Dinas Pariwisata Prov. Nusa Tenggara Barat pertahun yang didapat dari web Dinas Pariwisata NTB link : <http://www.dsbudpar.ntbprov.go.id/promosi-ntb/>.
3. Pengambilan dataset liburan sekolah untuk tiap semester dari tahun ajaran 2013/2014 semester 2 sampai dengan tahun pelajaran 2022/2023 semester 1 yang didapat dari website: <https://www.kalenderpendidikan.com/>

3.3. Metode Analisis Data

Dalam melakukan penelitian ini penulis menggunakan sistem analisa data kuantitatif dengan pendekatan metode analisis deskriptif. Nantinya hasil dari penelitian dengan menggunakan data historis jumlah wisatawan, data kalender *event* kegiatan, dan data libur sekolah. Dari variabel-variabel tersebut dilakukan prediksi untuk jumlah kedatangan wisatawan, nilai prediksi tersebut akan disajikan beserta dengan hasil nilai tingkat kesalahan (*error*) pada prediksi tersebut dengan pengukuran *MAPE*. *MAPE* sendiri adalah presentase rata-rata kesalahan secara mutlak (*absolut*). Menurut Khoiri (2020) *MAPE* adalah pengukuran statistika tentang akurasi perkiraan pada metode prediksi. Dimana apabila nilai tingkat kesalahan $<10\%$ maka kemampuan model dalam melakukan peramalan sangat baik, apabila hasil nilai kesalahan antara 10% - 20% maka kemampuan model dalam melakukan peramalan dianggap baik, sedangkan jika nilai tingkat kesalahan menghasilkan nilai tingkat kesalahan sebesar 20% - 50% maka kemampuan model dalam melakukan peramalan dianggap layak. Akan tetapi jika nilai tingkat kesalahan mencapai lebih dari 50% maka model dalam melakukan peramalan dianggap buruk. Jika hasil prediksi dan nilai kesalahan dengan menggunakan tiap variabel sudah didapatkan, maka akan dilakukan percobaan dengan melakukan kombinasi antar variabel, baik itu menggunakan satu, dua, atau tiga variabel secara bergantian dengan masing-masing metode yang digunakan. Berikut adalah alur analisa yang dilakukan.



Gambar 3.1. Metode analisis pengolahan data untuk prediksi

Dari alur gambar 3.1 diatas bisa diketahui bagaimana suatu data diolah untuk mendapatkan nilai prediksi menggunakan metode yang dipakai. Dari gambar 3.1 langkah dimulai dengan memasukan *dataset* kemudian pemodelan dan *preprocessing data* dan selanjutnya dilakukan perhitungan prediksi dengan metode *LSTM* dan *SVR* sehingga nilai prediksi untuk tiap variabel yang digunakan akan didapatkan. Hasil nilai prediksi selanjutnya dilakukan perhitungan kesalahan (*error*) untuk mengetahui seberapa besar tingkat kesalahan yang dilakukan oleh metode yang digunakan. Selanjutnya variabel-variabel yang lain akan dilakukan pengujian baik secara tunggal ataupun secara kombinasi untuk mendapatkan perbandingan nilai kesalahan yang paling akurat.

3.4. Alur Penelitian

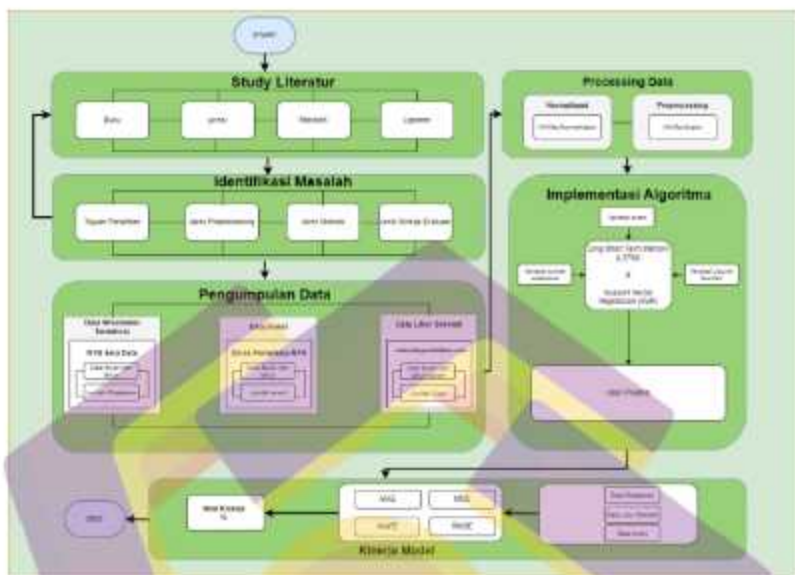
Dalam penelitian ini metode yang dilakukan mengacu kepada teknik observasi yang meliputi *identifikasi masalah* pada objek penelitian yang dilakukan dan serta studi pustaka yaitu dengan mengumpulkan serta menganalisa publikasi-publikasi yang telah dilakukan sebelumnya serta mengambil beberapa referensi teori dari beberapa buku yang telah terbit dan berkaitan dengan penelitian yang dilakukan. Selanjutnya adalah melakukan pengumpulan data wisatawan yang datang ke Provinsi Nusa Tenggara Barat dari tahun 2014 sampai dengan tahun 2022 yang diambil dari *dataset* publik di web resmi NTB Satu Data, data yang sudah

diambil kemudian dilakukan *preprocessing* data, dan kemudian dilakukan proses prediksi dengan menggunakan algoritma *Long Short Term Memory Network (LSTM)* dan juga *Support Vector Regression (SVR)*. Selanjutnya menguji kinerja algoritma terhadap data tersebut untuk mengetahui hasil akhir dari penelitian ini yaitu dengan melihat hasil tingkat kesalahan (*error*) dalam melakukan prediksi dengan menggunakan algoritma *LSTM* dan *SVR*. Setelah mengetahui hasil evaluasi kinerja algoritma, selanjutnya akan dilakukan *improvisasi* penggunaan kolaborasi antar variabel, sehingga nilai tingkat kesalahan bisa dibandingkan apakah variabel yang digunakan sudah berpengaruh terhadap prediksi atau malah tidak berdampak terhadap variabel yang lain. Adapun peta pemikiran penelitian tersebut dapat dilihat seperti pada gambar 3.2 berikut ini :



Gambar 3.2 Mind Map penelitian

Pada gambar 3.2 peta pemikiran (mind map) diatas dapat diketahui bahwa informasi mengenai penelitian ini dapat digambarkan dengan jelas sesuai dengan gambar 3.2 diatas, dimulai dari study literatur sampai ketahap pengujian prediksi digambarkan dengan jelas pada mind map gambar 3.2 diatas. Sedangkan untuk alur penelitian sendiri dapat dilihat dari gambar 3.3 dibawah ini :



Gambar 3.3 Alur Penelitian

Pada gambar 3.3 diatas bisa dilihat dengan jelas alur penelitian dimulai dari melakukan study literatur sampai mendapatkan hasil pengujian prediksi yang diharapkan.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

Hasil penelitian tentang pengujian variabel *event* dan libur sekolah yang telah dilakukan penulis sajikan pada bab keempat ini. Berikut adalah hasil pengujian serta pembahasan dari penelitian yang dilakukan.

4.1. Membangun Dataset

4.1.1. Pengumpulan Data

Dalam penelitian tentang pengujian variabel *event* dan libur sekolah data yang digunakan untuk melakukan prediksi ada tiga macam yaitu data jumlah kunjungan wisatawan dari tahun 2014 sampai dengan tahun 2022 di daerah Provinsi Nusa Tenggara Barat. Data jumlah kunjungan wisatawan tersebut merupakan data sekunder yang disediakan oleh database NTB satu data pada alamat <https://data.ntbprov.go.id/search/type/dataset>. Dataset yang disediakan merupakan data publik gratis yang dapat diunduh oleh setiap orang yang membutuhkan data tersebut guna kepentingan penelitian ataupun kepentingan lainnya. Data yang disediakan dari tahun 2014 sampai dengan tahun 2022 ini berjumlah 108 data. Data jumlah kunjungan wisatawan yang ada merupakan data yang terdiri dari dua kondisi (masa) yang berbeda, kondisi pertama adalah kondisi dimana wisatawan datang ke Provinsi NTB dalam keadaan (masa) normal tidak terjadi peristiwa istimewa yang terjadi pada masa tersebut, dan kondisi yang kedua merupakan kondisi (masa) dimana pada

waktu tersebut terjadi peristiwa *covid-19* dari tahun 2020 sampai dengan tahun 2021 sehingga pada kondisi (masa) tersebut kunjungan wisatawan yang datang ke Provinsi Nusa Tenggara Barat terjadi pengurangan wisatawan yang sangat drastis. Akan tetapi meskipun dataset yang digunakan merupakan dataset campuran dua kondisi dengan menggunakan metode *LSTM* dan *SVR* dataset tersebut bisa diolah dengan baik.

Selanjutnya dataset *event* kegiatan tahunan yang ada di Provinsi Nusa Tenggara Barat dari tahun 2017 sampai dengan tahun 2022. Data *event* ini juga merupakan data sekunder yang disediakan oleh Dinas Pariwisata Provinsi NTB yang didapat pada alamat <http://www.disbudpar.ntbprov.go.id/promosi-ntb/>. Selanjutnya dataset libur sekolah dari tahun pelajaran 2013/2014 semester 2 sampai dengan tahun pelajaran 2022/2023 semester 1 yang didapat dari website publik <https://kalenderpendidikan.com>.

Data yang telah dikumpulkan tersebut selanjutnya dikumpulkan kedalam aplikasi *microsoft excel* dengan format *.xlsx* file tersebut terdiri dari 4 kolom data yaitu kolom pertama berisi keterangan tanggal, bulan, dan tahun untuk menentukan deret waktu perbulan, kolom kedua berisi dataset jumlah kunjungan wisatawan, kolom ketiga berisi dataset *event* dan kolom keempat berisi dataset libur sekolah. Data yang telah dikumpulkan menjadi data file *.xlsx* selanjutnya siap dipakai untuk diproses oleh sistem pengolahan data.

4.1.2. Pengujian Dataset

Data yang telah disiapkan selanjutnya diupload kedalam *google colab* untuk selanjutnya dilakukan tahap pengolahan dataset. Data yang telah diupload akan dilakukan proses *preprocessing data*. Pada penelitian ini *preprocessing* yang digunakan adalah *min-max scaller* supaya data yang ada diubah kedalam data dengan range antara 0 sampai dengan 1, sehingga data yang ada bisa lebih mudah untuk dilakukan pengujian *training dataset* dan juga *testing dataset*. Pada kode dibawah ini adalah kode untuk melakukan upload dataset kedalam *google colab*.

```
import pandas as pd
import numpy as np
from datetime import datetime
from math import sqrt
from numpy import concatenate
from matplotlib import pyplot
from pandas import read_excel
from pandas import DataFrame
from pandas import concat
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import mean_squared_error
from keras.models import Sequential, load_model
from keras.layers import Dense
from keras.layers import LSTM

dataset = read_excel('/content/dataset wisatawan tesis new
3.xlsx', header=0, index_col=0)
dataset
```

Setelah *dataset diupload* kedalam sistem sesuai dengan perintah pada kode diatas selanjutnya *dataset* akan menampilkan hasil data yang telah diupload kedalam *google colab* seperti terlihat pada gambar 4.1 berikut.



tanggal	wisatawan	COE	libur
2017-01-01	148252	0	0
2017-02-01	162279	1	0
2017-03-01	158814	1	1
2017-04-01	221701	3	2
2017-05-01	235888	1	5
2017-06-01	306186	1	13
2017-07-01	422834	1	15

Gambar 4.1. dataset yang telah diupload kedalam google colab

Hasil dari pemanggilan dataset yang telah diupload kedalam google colab dapat terlihat pada tabel 4.2 dengan jumlah kolom adalah empat kolom dengan kolom pertama berisi tanggal untuk deret waktu, kolom kedua berisi jumlah wisatawan, kolom ketiga berisi data *event* dan kolom terakhir berisi data libur sekolah.

4.1.3. Preprocessing Data

Setelah dataset berhasil diupload kedalam google colab seperti yang terlihat pada gambar 4.1 selanjutnya data akan dilakukan *preprocessing*. Berikut adalah penerapan *minmaxscaller* untuk melakukan *preprocessing data* dapat dilihat pada gambar 4.2.

```
scaler = MinMaxScaler(feature_range=(0, 1))
scaled = scaler.fit_transform(values)
scaled
```

```
array([[0.25839674, 0.00000000, 0.28660556],
       [0.28660556, 0.16666667, 0.27923444],
       [0.27923444, 0.16666667, 0.40610000],
       [0.40610000, 0.50000000, 0.43463200],
       [0.43463200, 0.16666667, 0.22727273],
       [0.57599563, 0.16666667, 0.59090909],
       [0.8105712, 0.16666667, 0.68181818],
       [0.88087774, 0.33333334, 0.04545455],
       [0.9504917, 0.83333334, 0.09090909],
       [0.54619384, 0.33333334, 0.00000000],
       [0.5131336, 0.83333334, 0.00000000],
       [0.5004283, 0.33333334, 0.68181818],
       [0.35060954, 0.16666667, 0.04545455],
       [0.39090567, 0.33333334, 0.04545455],
       [0.3808725, 0.16666667, 0.09090909],
       [0.0788193, 0.33333334, 0.04545455],
       [0.0034906, 0.50000000, 0.13636364],
       [0.8424751, 0.16666667, 1.00000000],
       [0.99999994, 0.16666667, 0.09090909]])
```

Gambar 4.2 hasil *preprocessing minmax scaler*

Pada gambar 4.2 diatas data dinormalisasi dengan melakukan konversi menjadi rentang nilai antara 0 dan 1. Setelah data dilakukan normalisasi selanjutnya dilakukan pemilihan variabel yang akan digunakan dalam memprediksi dimulai penggunaan variabel histori wisatawan dan *event* untuk prediksi, variabel histori wisatawan dan libur sekolah serta variabel histori wisatawan, *event* dan libur sekolah yang digunakan secara bersamaan.

	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 4.3 pemilihan variabel metode *LSTM* var 1 dan 2

Dari gambar 4.3 diatas data yang digunakan dalam melakukan prediksi wisatawan dengan metode *LSTM* hanya menggunakan variabel histori wisatawan yaitu *var1*, dan juga variabel even 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 4.4 pemilihan variabel metode LSTM variabel 1 dan 3

Dari gambar 4.4 diatas data yang digunakan dalam melakukan prediksi wisatawan dengan metode LSTM 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 4.5 pemilihan variabel metode LSTM semua variabel

Sedangkan pada gambar 4.5 diatas data yang digunakan dalam melakukan prediksi wisatawan dengan metode LSTM menggunakan semua variabel yang digunakan yaitu histori wisatawan sebagai *var1*, variabel even sebagai *var2* dan variabel liburan sekolah sebagai *var3*. Ketiga pemilihan variabel tersebut digunakan untuk melihat kinerja pada masing-masing variabel pada dataset yang digunakan.

```
x = data[['Wisatawan', 'Event', 'Libur']]
y = data['Wisatawan']
x
```



	wisatawan	libur	COE
0	108551	8	0
1	97095	0	1
2	105129	1	1
3	124233	2	3
4	142279	5	1
...	...	...	...
103	127042	1	6
104	121794	0	2
105	125205	1	4
106	129427	0	6
107	154209	15	4

108 rows x 3 columns

Gambar 4.6 pemilihan variabel metode SVR semua variabel

Pada gambar 4.6 adalah pemilihan fitur/variabel yang akan digunakan dalam melakukan prediksi menggunakan metode SVR. Pada gambar 4.6 tersebut variabel yang digunakan adalah semua variabel yaitu data wisatawan, libur dan data event

```
x = data[['Wisatawan', 'Libur']]
y = data['Wisatawan']
x
```



	wisatawan	libur
0	108551	8
1	97096	0
2	105129	1
3	124233	2
4	142279	5
...	...	...
103	127042	1
104	121794	0
105	125265	1
106	129427	0
107	154209	15

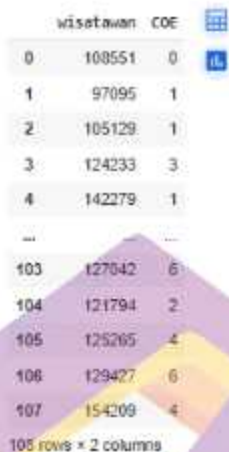
108 rows x 2 columns

Gambar 4.7 pemilihan variabel metode *SIR* wisatawan dan libur

4.2. Membangun Model

Selanjutnya pada gambar 4.7 fitur/variabel yang digunakan dalam melakukan prediksi wisatawan menggunakan metode *SIR* menggunakan dua variabel yaitu variabel wisatawan dan libur sebagai variabel independennya dan variabel wisatawan sebagai variabel dependennya.

```
x = data[['Wisatawan', 'COE']]
y = data['Wisatawan']
x
```



	wisatawan	COE
0	108551	0
1	97095	1
2	105129	1
3	124233	3
4	142279	1
...	...	...
103	127042	6
104	121794	2
105	125265	4
106	129427	6
107	154209	4

105 rows x 2 columns

Gambar 4.8 pemilihan var metode *SVR* wisatawan dan *event*

Pada gambar 4.8 *fitur/variabel* yang digunakan dalam melakukan prediksi wisatawan menggunakan metode *SVR* adalah variabel wisatawan dan juga *event*.

4.2.1 Training dataset

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*, pembagian data *training* dan data *testing* menggunakan 80% data *training* dan 20% data *testing* (Josua Manulang Dkk, 2020) adapun hasil pelatihan model yang telah dilakukan adalah seperti berikut.

```
history = model.fit(train_X, train_y, epochs=200,
                    batch_size=32, validation_data=(test_X, test_y), verbose=2,
                    shuffle=False)
model.save('./savedModel')
```

```

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

```

Gambar 4.9. Melatih Model Neuron 20 Epoch 200 Batch Size 32

Dari gambar 4.9 dataset yang digunakan dalam melakukan prediksi dibagi menjadi *data training* dan *data testing* dengan jumlah pembagian masing-masing data adalah 80/20. Dari gambar 4.9 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. Adapun pemilihan *data splitting* menggunakan skenario 80/20 karena skenario tersebut merupakan skenario paling banyak digunakan oleh peneliti dalam melakukan *splitting data* (Josua Manulang Dkk, 2020). Selanjutnya kita bandingkan model dengan jumlah *neuron* 40, *epoch* 150 dan *batch size* 16 dapat dilihat pada gambar 8 dibawah ini.

```

history = model.fit(train_X, train_y, epochs=150,
batch_size=16, validation_data=(test_X, test_y), verbose=2,
shuffle=False)
model.save('./savedModel')

```

```

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 4.10. Melatih Model *Neuron 40 Epoch 200 Batch Size 32*

Pada gambar 4.10 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 4.11 dibawah dapat dilihat *hyperparameter* terbaik untuk menghasilkan kinerja model.

```

estimator = KerasRegressor(build_fn=create_model)
random_search = RandomizedSearchCV(estimator,
param_distributions=param_dist, cv=3, n_iter=10,
scoring='neg_mean_squared_error')
random_search.fit(train_X, train_y)
print('Best hyperparameters: ', random_search.best_params_)

```

Pada kode diatas metode pencarian parameter terbaik menggunakan *randomizedSearch*, dengan hasil yang diperoleh dari penggunaan parameter diatas adalah jumlah *neuron 40*, jumlah *epoch 200*, *dropout 0.5* dan *batch_size 16* (Jian-Wu Bi, 2020). Berbeda dengan metode LSTM, pada metode SVR pencarian parameter terbaik menggunakan metode *gridsearchCV*, dengan parameter yang akan dicari adalah nilai *C*, *epsilon*, dan *kernel*. Hasil

yang diperoleh dari pencarian *gridsearchcv* dapat dilihat pada gambar 4.11 dibawah ini.

```
Best parameters: {'C': 0.1, 'epsilon': 0.1, 'kernel': 'linear'}
Best score: 13358.365815599454
```

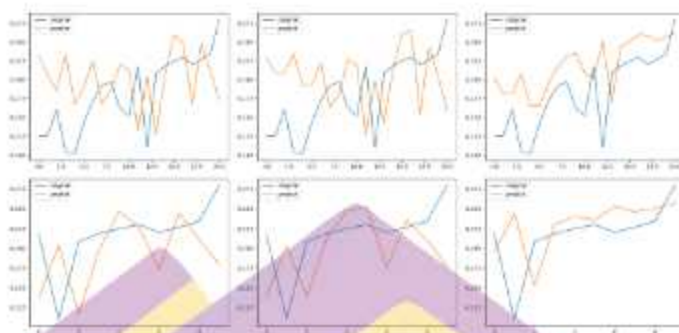
Gambar 4.11 hasil pencarian parameter terbaik SVR

Dari gambar 4.11 diatas hasil pencarian paramter terbaik menggunakan metode *gridsearchcv* adalah nilai *C* 0.1 nilai *epsilon* 0.1 dan *kernel* yang digunakan adalah *kernel linear*.

4.2.2 Pengujian Metode LSTM

Setelah model dilatih tahapan selanjutnya adalah pengujian variabel menggunakan metode LSTM dan SVR. 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 *independent* yaitu variabel *event* dan libur sekolah serta variabel *dependent* yaitu jumlah wisatawan.

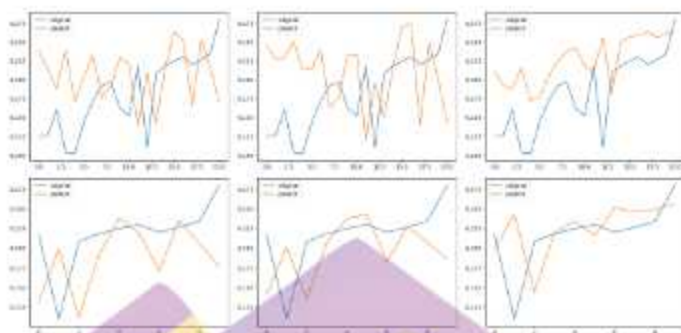
Untuk pengujian prediksi metode LSTM tiap variabel yaitu pengujian pertama menggunakan variabel histori, *event* dan liburan, pengujian kedua histori wisatawan dan *event* saja, serta pengujian ketiga menggunakan variabel histori wisatawan dan liburan saja disajikan berdasarkan dua parameter pengujian seperti terlihat pada gambar grafik berikut yaitu gambar 4.15 menggunakan pengujian dengan parameter kesatu dan pada gambar 4.16 merupakan pengujian dengan paramater kedua.



Gambar 4.12. Prediksi parameter *Neuron* : 40, *Epoch* :150, *Batch_size* : 16

Pada gambar 4.12 diatas untuk bagian gambar yang paling atas hasil pengujian dengan data *split 70/30* adapun untuk atas yang pertama adalah pengujian semua variabel, dan untuk atas yang kedua adalah pengujian variabel *event*, Sedangkan pada bagian atas yang ketiga adalah hasil pengujian variabel libur sekolah. Sedangkan bagian bawah adalah pembagian data *splitting 80/20* untuk bagian bawah yang pertama pengujian variabel semua variabel, untuk bagian bawah kedua adalah hasil pengujian untuk variabel *event*, dan untuk bagian bawah yang ketiga adalah pengujian variabel libur sekolah.

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 4.16 dibawah ini.



Gambar 4.13. Prediksi parameter *Neuron* : 20, *Epoch* :200, *Batch_size* : 32

Pada gambar 4.13 posisi pengujian sama persis dengan yang ada pada gambar 4.12. Setelah itu kemudian dilakukan pengujian untuk melihat hasil prediksi dan data aktual, hasil pengujian menggunakan metode LSTM dapat dilihat pada tabel 4.1 dibawah ini.

Tabel 4.1. Hasil prediksi Parameter 1 semua variabel

No	Variabel	Prediksi		Aktual
		Splitting 70/30	Splitting 80/20	
1	Semua variabel	74355.67	76314.42	120697
2	Semua variabel	113184.57	113327.47	62083
3	Semua variabel	71819.74	65816.03	115324
4	Semua variabel	114574.6	109010.6	120764
5	Semua variabel	143006.88	135967.06	124331
6	Semua variabel	137155.	125125.35	127042
7	Semua variabel	93519.33	96682.15	121794.
8	Semua variabel	136993.88	135035.67	125265.
9	Semua variabel	116647.21	116753.664	129427.
10	Semua variabel	96058.05	99707.31	154209.

Data pada tabel 4.1 diatas menunjukkan hasil pengujian semua variabel dengan menggunakan *Neuron* : 20, *Epoch* :200, *Batch_size* : 32. Adapun untuk hasil pengujian variabel libur sekolah menggunakan parameter yang sama dapat dilihat pada tabel 4.2 berikut.

Tabel 4.2. Hasil pengujian variabel libur parameter 1

No	Variabel	Prediksi		Aktual
		<i>Splitting 70/30</i>	<i>Splitting 80/20</i>	
1	Liburan	111915.91	108069.055	120697
2	Liburan	138037.89	134783.61	62083
3	Liburan	94452.19	85172.46	115324
4	Liburan	134660.16	127078.64	120764
5	Liburan	138396.94	132659.61	124331
6	Liburan	143281.34	130400.945	127042
7	Liburan	142676.88	139530.1	121794
8	Liburan	138821.7	136022.95	125265
9	Liburan	141377.48	138200.53	129427
10	Liburan	144377.22	141704.66	154209

Pada tabel 4.2 pengujian variabel libur sekolah menunjukan antara nilai prediksi dengan nilai sebenarnya. Selanjutnya pada tabel 4.3 dibawah ini adalah hasil dari pengujian variabel *event* menggunakan parameter 1.

Tabel 4.3 pengujian variabel *event* dengan parameter 1

No	Variabel	Prediksi		Aktual
		<i>Splitting 70/30</i>	<i>Splitting 80/20</i>	
1	<i>Event</i>	70935.22	77260.125	120697
2	<i>Event</i>	109778.67	112554.95	62083
3	<i>Event</i>	79591.96	77681.62	115324
4	<i>Event</i>	118183.266	117256.87	120764
5	<i>Event</i>	143747.56	136389.67	124331
6	<i>Event</i>	146304.94	139139.9	127042
7	<i>Event</i>	86964.37	97429.766	121794
8	<i>Event</i>	134261.4	130053.87	125265
9	<i>Event</i>	113102.31	115925.484	129427
10	<i>Event</i>	88666.15	99130.945	154209

Dari tabel 4.3 diatas variabel yang diuji adalah variabel *event* menghasilkan nilai prediksi yang berbeda antara pengujian menggunakan *splitting 70/30* dan juga *80/20* terhadap nilai aktual. Dari hasil pengujian yang dilakukan berdasarkan tiga skenario pengujian selanjutnya dilakukan pengamatan perbandingan antar variabel yang digunakan dalam melakukan prediksi menggunakan pengujian parameter 1, hasil pengamatan analisis perbandingan variabel dapat dilihat pada tabel 4.4 dibawah ini.

Tabel 4.4. perbandingan nilai pengujian

No	Variabel				Aktual
	<i>Splitting</i>	<i>Event</i>	Libur Sekolah	Gabungan variabel	
1	70/30	70935.22	111915.91	74355.67	120697
2	80/20	77260.125	108069.055	76314.42	

Dari tabel 4.4 diatas hasil pengujian prediksi dari variabel *event* memiliki rentang nilai prediksi sebesar 49761,78 untuk *splitting* 70/30 dan 43436,875 untuk *splitting* 80/20. Sedangkan untuk variabel libur sekolah memiliki rentang nilai prediksi sebesar 8781,09 untuk *splitting* 70/30 dan 12627,945 untuk *splitting* 80/20. Sedangkan hasil dari variabel gabungan menghasilkan rentang nilai prediksi sebesar 46341,33 untuk *splitting* 70/30 dan 44382,58 untuk *splitting* 80/20. Dari pengujian menggunakan parameter 1 didapatkan variabel dengan rentang yang mendekati nilai aktual adalah variabel libur sekolah dengan *splitting* 70/30.

Selanjutnya setelah pengujian parameter 1 selesai berikut adalah hasil prediksi pengujian untuk parameter yang kedua yaitu jumlah *neuron* 40, *epoch* 150 dan *batch size* 16 dengan metode LSTM. Hasil pengujian semua variabel menggunakan parameter kedua dengan data *splitting* 70/30 dan juga 80/20 dan juga data aktual bisa dilihat pada tabel 4.5 dibawah ini.

Tabel 4.5 Hasil prediksi parameter 2 semua variabel

No	Variabel	Prediksi		Aktual
		<i>Splitting</i> 70/30	<i>Splitting</i> 80/20	
1	Semua variabel	77311.46	73506.734	120697
2	Semua variabel	115924.016	110659.54	62083
3	Semua variabel	79241.266	63454.1	115324
4	Semua variabel	119553.6	106216.516	120764
5	Semua variabel	145167.8	131443.92	124331
6	Semua variabel	138962.1	121570.38	127042
7	Semua variabel	92043.945	95262.34	121794.
8	Semua variabel	139676.92	130162.234	125265.
9	Semua variabel	119139.21	114157.125	129427.
10	Semua variabel	94145.15	98363.47	154209.

Pada tabel 4.5 diatas adalah hasil dari pengujian semua variabel menggunakan parameter kedua dengan *data splitting* 70/30 dan juga 80/20. Selanjutnya pengujian variabel libur sekolah menggunakan parameter kedua dengan *data splitting* 70/30 dan 80/20 dapat dilihat di tabel 4.6 dibawah ini.

Tabel 4.6 Hasil pengujian variabel libur sekolah parameter 2

No	Variabel	Prediksi		Aktual
		<i>Splitting</i> 70/30	<i>Splitting</i> 80/20	
1	Libur sekolah	116514.516	109313.945	120697
2	Libur sekolah	140837.17	134181.84	62083
3	Libur sekolah	100899.234	80811.37	115324
4	Libur sekolah	137948.11	121755.84	120764
5	Libur sekolah	141744.25	129106.46	124331
6	Libur sekolah	142550.97	119576.19	127042
7	Libur sekolah	145193.12	138652.45	121794
8	Libur sekolah	141294.69	135928.39	125265
9	Libur sekolah	143971.92	137398.64	129427
10	Libur sekolah	146518.62	141259.02	154209

Dari tabel 4.6 diatas hasil pengujian variabel libur sekolah untuk *splitting* 70/30 dan 80/20 serta data aktual disajikan pada tabel 4.5 diatas. Selanjutnya pengujian variabel *event* menggunakan parameter 2 dengan metode LSTM dengan *splitting* 70/30 dan 80/20 dapat dilihat pada tabel 4.7 berikut.

Tabel 4.7. Hasil pengujian variabel *event* Parameter 2

No	Variabel	Prediksi		Aktual
		<i>Splitting</i> 70/30	<i>Splitting</i> 80/20	
1	<i>Event</i>	67264.305	78278.79	120697
2	<i>Event</i>	108112.41	111640.89	62083
3	<i>Event</i>	84089.39	75947.1	115324
4	<i>Event</i>	119549.41	114605.25	120764
5	<i>Event</i>	148679.39	131310.9	124331
6	<i>Event</i>	151075.69	133951.11	127042
7	<i>Event</i>	77976.27	101402.24	121794
8	<i>Event</i>	137487.06	125951.53	125265
9	<i>Event</i>	111254.75	115029.42	129427
10	<i>Event</i>	79628.305	103158.66	154209

Tabel 4.7 diatas adalah pengujian terakhir dari parameter kedua yaitu pengujian variabel *event* dengan hasil prediksi berdasarkan *data splitting* 70/30

dan 80/20. Dari hasil pengujian yang dilakukan berdasarkan tiga skenario pengujian selanjutnya dilakukan pengamatan perbandingan antar variabel yang digunakan dalam melakukan prediksi menggunakan pengujian parameter 2, hasil pengamatan analisis perbandingan variabel dapat dilihat pada tabel 4.8 dibawah ini.

Tabel 4.8. perbandingan nilai pengujian parameter 2

No	Variabel				Aktual
	<i>Splitting</i>	<i>Event</i>	Libur Sekolah	Gabungan variabel	
1	70/30	67264.305	116514.516	77311.46	120697
2	80/20	78278.79	109313.945	73506.734	

Dari tabel 4.8 diatas hasil pengujian prediksi dari variabel *event* memiliki rentang nilai prediksi sebesar 53432,695 untuk *splitting* 70/30 dan 42418,21 untuk *splitting* 80/20. Sedangkan untuk variabel libur sekolah memiliki rentang nilai prediksi sebesar 4182,484 untuk *splitting* 70/30 dan 11383,055 untuk *splitting* 80/20. Sedangkan hasil dari variabel gabungan menghasilkan rentang nilai prediksi sebesar 43385,54 untuk *splitting* 70/30 dan 47190,266 untuk *splitting* 80/20. Dari pengujian menggunakan parameter 1 didapatkan variabel dengan rentang yang mendekati nilai aktual adalah variabel libur sekolah dengan *splitting* 70/30.

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*. Sedangkan untuk pengukuran utama yang digunakan oleh penulis disini adalah pengukuran *MAPE*, penulis lebih memilih pengukuran *MAPE* dibandingkan dengan pengukuran yang lain karena pengukuran *MAPE* merupakan satuan

nilai pengukuran yang digunakan untuk mengukur tingkat efektifitas prediksi suatu metode dengan melihat rata-rata tingkat error dari metode yang diuji. Sehingga pembacaan hasilnya pun lebih mudah dipahami oleh pembaca karena hasil pengukuran disajikan berupa nilai persentase (%). Semakin kecil nilai MAPE maka semakin baik pula model yang dihasilkan dalam melakukan prediksi. Hasil pengukuran dapat dilihat pada tabel 4.9 berikut ini.

Tabel 4.9. Pengujian parameter *Neuron : 40, Epoch :150, Batch_size : 16*

No	Variabel	Splitting 70/30			Splitting 80/20		
		MAPE	RMSE	MAE	MAPE	RMSE	MAE
1	Libur	106.3	50537.6	45652	30.343	29554.2	24218.8
2	Event	110.5	57261.7	51632	38.981	37982.2	33232.1
3	Semua Variabel	103.7	53989.1	48793	36.25	35913.9	30752.1

Pada tabel 4.9 diatas nilai MAPE terkecil didapatkan pada pengujian liburan disemua pengukuran kesalahan (*error*) dengan nilai MAPE sebesar 30,343% dengan data training 80% dan data testing 20%. Sedangkan dengan *splitting* 70% data training dan 30% data testing pengujian terbaik ada pada pengujian semua variabel untuk pengukuran MAPE dengan nilai MAPE sebesar 103.7% untuk pengukuran RMSE dan MAE hasil terbaik pada pengujian libur sekolah dengan nilai RMSE sebesar 50537.6 dan nilai MAE sebesar 45652. Sedangkan pada tabel 4.10 dibawah adalah hasil pengujian kinerja menggunakan parameter kedua yaitu dengan jumlah *neuron* 20, *epoch* 200 dan *batch size* 32.

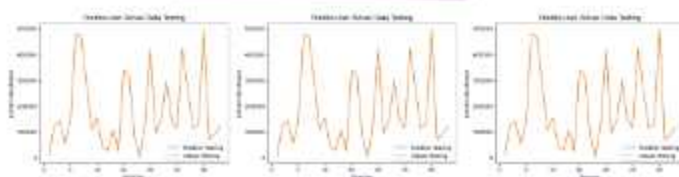
Tabel 4.10. Pengujian parameter *Neuron : 20, Epoch :200, Batch_size : 32*

No	Variabel	Splitting 70/30			Splitting 80/20		
		MAPE	RMSE	MAE	MAPE	RMSE	MAE
1	Libur	108	50807.4	45646.8	33.503	32234.5	26672
2	Event	104	52786	47362	45.345	43874.3	38624.1
3	Semua Variabel	104	52507.1	47196	37.911	37492.6	31786.1

Berdasarkan hasil pengujian dengan menggunakan berbagai parameter 2 sesuai pada tabel 4.10 didapatkan hasil terbaik menggunakan data *splitting* 80% untuk *training* dan 20% untuk *testing* dalam pengujian semua pengukuran kesalahan (*error*) dengan hasil pengukuran MAPE sebesar 33.503% nilai RMSE sebesar 32234.5 dan nilai MAE sebesar 26672. Sedangkan pengujian *splitting* 70/30 hasil terbaik dengan pengukuran MAPE pada variabel *event* dan semua variabel pengujian dengan nilai kinerja kesalahan MAPE sebesar 104% sedangkan untuk pengukuran RMSE dan MAE hasil terbaik pada variabel libur sekolah dengan nilai RMSE sebesar 50807.4 dan nilai MAE sebesar 45646.8. Dari kedua parameter yang digunakan hasil terbaik didapat pada parameter pertama 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*.

4.2.3 Pengujian Metode Support Vector Regression (SVR)

Hasil pengujian pertama dilakukan untuk menguji variabel wisatawan, *event* dan juga variabel libur dengan skema 1 dengan penggunaan *splitting* 70/30 disajikan berupa grafik yang dapat dilihat pada gambar 4.17 berikut.



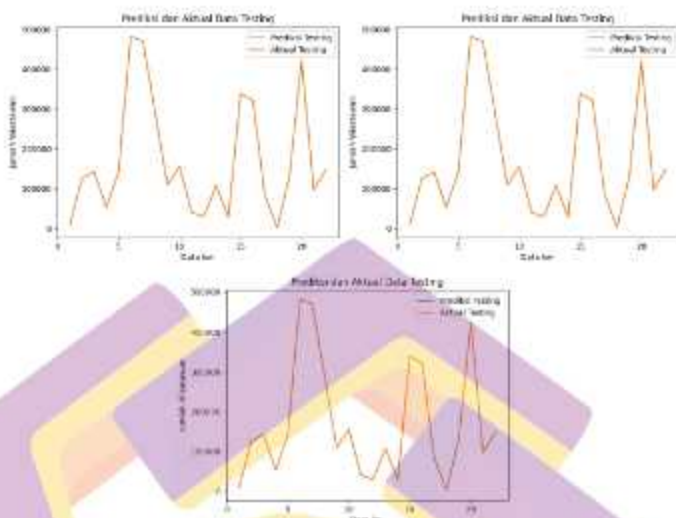
Gambar 4.14. Pengujian *splitting* 70/30

Dari grafik gambar 4.14 dengan grafik bagian atas sebelah kiri adalah hasil pengujian variabel wisatawan dan libur sekolah, grafik bagian atas sebelah kanan adalah hasil pengujian variabel wisatawan dan *data event* sedangkan grafik sebelah bawah adalah hasil pengujian semua variabel, dari gambar 4.14 tampak jelas bahwa prediksi sangat mendekati terhadap data aktual yang berarti model sangat baik dalam melakukan prediksi. Adapun untuk data hasil prediksi dan juga data aktual dapat dilihat pada tabel 4.11 berikut.

Tabel 4.11. Pengujian splitting 70/30

No	Data- Ke	Prediksi			Aktual
		Wisatawan-libur	Wisatawan-event	Semua data	
1	77	9015.88037555	9240.80585021	9074.86925405	9085
2	10	126122.06000361	126099.01690282	126049.90859055	126074
3	4	142323.36591944	142278.98430194	142211.70261264	142279
4	83	53498.58591554	53642.79982199	53491.92823243	53539
5	62	141615.06883583	141561.34527697	141516.54086947	141559
6	67	481457.69380269	480764.05263938	480879.78029418	481159
7	30	470375.10367093	469740.44365539	469774.59804511	470124
8	45	291532.1517857	291193.08845653	291225.04801917	291366
9	95	109352.67060051	109381.03660526	109279.3607676	109341
10	11	156111.68838677	156073.34803619	155952.12827754	156088

Pada tabel 4.11 prediksi sangat mendekati terhadap nilai aktual, data yang digunakan menggunakan data acak dengan jumlah prediksi sebanyak 10 data. Selanjutnya pengujian variabel wisatawan dan libur sekolah menggunakan metode SVR dengan *data splitting 80/20* dapat dilihat pada gambar 4.18 dibawah ini.



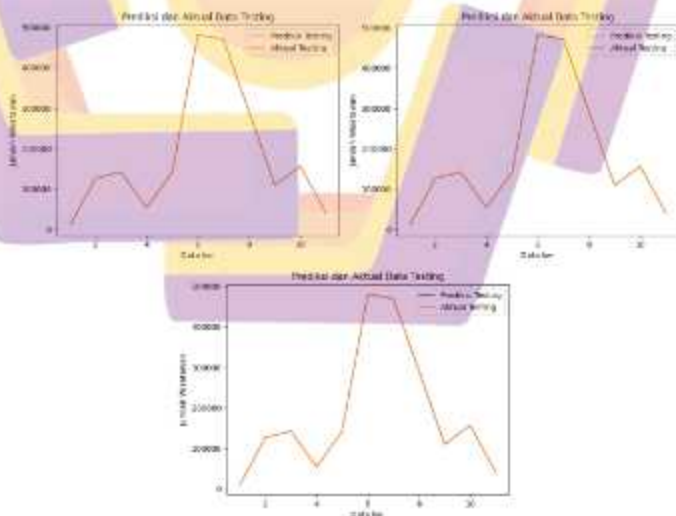
Gambar 4.15 pengujian SVR dengan *data splitting* 80/20

Pada gambar 4.15 bagian atas sebelah kiri adalah hasil pengujian variabel wisatawan dan libur sekolah, grafik bagian atas sebelah kanan adalah hasil pengujian variabel wisatawan dan *data event* sedangkan grafik sebelah bawah adalah hasil pengujian semua variabel, dengan pengujian menggunakan metode SVR dan dengan *data splitting* 80/20 juga terlihat bahwa data prediksi persis mengikuti pola data aktual hal ini menunjukkan bahwa pembagian *data training* dan *data testing* pada skema 2 juga sangat baik dalam membaca pola pada model yang dibuat. Tabel hasil pengujian prediksi terhadap data aktual dapat dilihat pada tabel 4.12 dibawah ini.

Tabel 4.12 pengujian *splitting* 80/20

No	Data Ke-	Prediksi			Aktual
		Wisatawan-libur	Wisatawan-event	Semua data	
1	77	9007.10296283	9067.08967164	9042.60756487	9085
2	10	126093.22328427	126080.32179971	126091.10572273	126074
3	4	142339.92368939	142276.57697196	142332.82319242	142279
4	83	53520.43581643	53526.25861926	53538.80049497	53539
5	62	141601.71285798	141559.1505291	141597.32124883	141559
6	67	481621.89931204	481198.63796992	481493.99999219	481159
7	30	470629.53567984	470159.69761844	470494.4489401	470124
8	45	291589.73878004	291383.56952948	291531.83709699	291366
9	95	109374.15954091	109337.40433497	109374.67687828	109341
10	11	156209.30950072	156089.83990283	156186.01726991	156088

Pada tabel 4.12 selisih antara data hasil prediksi dengan data aktual sangat kecil sehingga dari grafik pada gambar 6 dengan data hasil pengujian pada tabel 4.7 sudah sesuai dan sudah sangat baik dalam melakukan prediksi. Selanjutnya pengujian dengan menggunakan skema ketiga dengan pembagian data *split* 90/10 dapat dilihat pada gambar 4.19 dibawah ini.

Gambar 4.16. Prediksi *splitting* 90/10

Pada gambar 4.16 grafik yang berada dibagian atas sebelah kiri adalah hasil pengujian menggunakan variabel wisatawan dan libur sekolah, grafik bagian atas sebelah kanan adalah hasil pengujian variabel wisatawan dan *data event* sedangkan grafik sebelah bawah adalah hasil pengujian semua variabel, model yang dibuat menggunakan metode SVR ini sangat persis mengikuti model yang telah dibuat yang menunjukkan tingkat error model yang dibuat sangat kecil, hal ini berarti metode yang digunakan dalam melakukan prediksi sudah sangat baik.

Tabel 4.13. Pengujian *splitting* 90/10

No	Data-Ke	Prediksi			Aktual
		Wisatawan-libur	Wisatawan-event	Semua data	
1	77	9069.28602674	9246.50404072	9033.22932193	9085
2	10	126061.35435225	126006.82475748	126195.47866336	126074
3	4	142231.5175965	142185.47731628	142403.84112588	142279
4	83	534964.5400084	53615.38790812	53539.9796671	53539
5	62	141533.4853902	141465.68583502	141692.78160134	141559
6	67	480969.7981136	480415.45330562	481711.52113953	481159
7	30	469869.84956051	469402.75208988	470634.51833192	470124
8	45	291273.68354489	290985.85018386	291686.5623211	291366
9	95	109293.17817642	109309.37367859	109420.61388007	109341
10	11	155979.80978657	155966.86713352	156210.81295911	156088

Pada tabel 4.13 dapat dilihat perbandingan selisih didapatkan paling kecil dibanding dengan pengujian skema *data splitting* 70/30 pada tabel 4.11 dan pengujian skema 2 *data splitting* 80/20 pada tabel 4.12.

Setelah pengujian prediksi terhadap data aktual menggunakan metode SVR selesai selanjutnya data hasil uji dilakukan pengujian untuk melihat nilai kinerja evaluasi, pengujian ini menggunakan evaluasi kinerja *MAE*, *MAPE* dan

RMSE. Adapun hasil beberapa pengujian dengan skema yang telah ditetapkan dapat dilihat pada tabel 4.7 dibawah ini.

Tabel 4.14. Hasil pengujian kinerja

Dataset	Splitting 70/30			Splitting 80/20			Splitting 90/10		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Event	103803.338	42.69626	156801.6167	117572	41.3125	174883.89	112134.44	48.15867	170211.87
Libur sekolah	103803.101	42.69623	156801.2877	117571.68	41.3124	174883.49	112134.15	48.15861	170211.58
Semua dataset	103803.107	42.69624	156801.2811	117571.67	41.3125	174883.45	112134.14	48.15859	170211.59

Dari tabel 4.14 dapat dilihat bahwa pengujian dilakukan dengan 3 skema yang berbeda berdasarkan *splitting* 70/30, 80/20 dan 90/10 dan menggunakan tiga evaluasi kinerja yaitu MAE, MAPE, dan RMSE. Hasil pengujian terbaik tiap-tiap skema *splitting* dapat dilihat pada tabel 4.14 dengan diberi warna pada *cell* tabel, pengujian terbaik menggunakan *splitting* 70/30 menggunakan pengukuran MAPE berada pada pengujian variabel libur sekolah dengan nilai error 42.69623%, pengujian terbaik menggunakan *splitting* 80/20 pada variabel libur sekolah dengan nilai error 41.3124%, sedangkan pengujian *splitting* 90/10 hasil pengujian terbaik pada variabel semua variabel dengan nilai error 48.15859%. secara keseluruhan hasil terbaik pada pengujian *splitting* 80/20 variabel libur sekolah dengan nilai error didapat adalah 41.3124%. Hasil dari ketiga parameter menunjukan bahwa ketiga parameter tersebut menghasilkan pengujian variabel *event* dan libur sekolah dapat dengan baik digunakan untuk melakukan prediksi terhadap wisatawan.

Dari hasil pengujian kedua metode yang dilakukan ada beberapa penelitian yang dijadikan referensi oleh penulis yang menghasilkan pengujian dengan hasil yang sama seperti yang telah penulis lakukan, pengujian pada penelitian ini sejalan dengan penelitian yang dilakukan oleh Rui Zhang dkk bahwa LSTM kurang baik jika menggunakan dataset yang terbatas, akan tetapi jika dataset yang digunakan ditingkatkan lagi jumlahnya maka masalah seperti *overfitting* bisa diatasi. Penelitian yang dilakukan oleh Ghahreman dkk, 2020, J. Gourav Bathla, 2020, juga sejalan dengan hasil penelitian yang dilakukan pada penelitian ini yaitu hasil dari LSTM tampil lebih baik dibanding dengan metode lainnya sebagai pembandingan.

Penelitian yang dilakukan oleh Ahmad Kamal Badri dkk, 2022 juga sejalan dengan hasil penelitian yang dilakukan oleh penulis, pada penelitian yang dilakukan oleh Ahmad Kamal Badri dkk menyatakan bahwa pengujian menggunakan dataset dua kondisi yaitu ketika keadaan normal dan ketika covid-19 bisa dilakukan dengan baik oleh metode LSTM, hal itu sejalan dengan apa yang telah peneliti lakukan pada penelitian ini.

Sedangkan penelitian yang dilakukan oleh Yogi firgiawan dalam melakukan prediksi tidak sama dengan hasil yang didapat pada penelitian yang peneliti hasilkan, pada penelitian Yogi firgiawan metode SVR tampil lebih unggul dalam semua pengujian dibanding dengan metode LSTM. Hal ini tidak sejalan dengan hasil penelitian yang dilakukan oleh peneliti dimana hasil menunjukan bahwa metode LSTM tampil lebih baik dibanding dengan metode SVR.

BAB V

PENUTUP

5.1. Kesimpulan

Dari hasil analisis yang telah dilakukan dalam melakukan penelitian ini, diperoleh beberapa kesimpulan dalam menguji variabel *event* dan libur sekolah dalam melakukan prediksi terhadap wisatawan, adapun kesimpulan yang didapat adalah sebagai berikut :

1. Pengujian dengan metode *LSTM* menghasilkan nilai kinerja lebih baik dibanding dengan metode *SVR* dalam melakukan prediksi wisatawan ketika data *splitting* yang digunakan 80/20 dengan nilai *MAPE* . Penggunaan data *splitting* menghasilkan nilai kinerja terbaik ini selaras dengan penelitian yang telah dilakukan oleh (Josua Manulang Dkk, 2020). Akan tetapi bertolak belakang dengan penelitian yang dilakukan oleh Yogi Firgiawan yang menghasilkan metode *SVR* tampil lebih baik dibanding dengan metode *LSTM*.
2. Metode *LSTM* dapat memberikan nilai kinerja yang tinggi menggunakan *hyperparameter Neuron : 40, Epoch :150, Batch_size : 16* dibanding dengan *hyperparameter* lainnya dengan nilai kinerja *MAPE* sebesar 30.343% dan data *splitting* 80/20 dengan dataset dua kondisi baik ketika sedang tidak ada kasus *covid-19* dan juga ketika sedang ada kasus *covid-19*. Hasil dengan menggunakan dataset kondisi normal saja tanpa melibatkan data ketika terjadi *covid-19* (2020-2021) malah menghasilkan kinerja lebih rendah dengan tingkat

error MAPE mencapai 56.27%. Hal ini selaras dengan penelitian yang dilakukan oleh (Rui Zhang dkk, 2022) yang menyatakan bahwa semakin banyak dataset yang digunakan maka hasil kinerja metode LSTM pun semakin baik. Dan selaras juga dengan penelitian yang dilakukan oleh (Ahmad Kamal Badri dkk, 2022) yang menunjukkan bahwa dataset dua kondisi bisa dengan baik digunakan untuk pengujian prediksi.

3. Variabel paling baik dalam melakukan prediksi wisatawan menggunakan metode LSTM adalah dengan pengujian variabel individual libur sekolah dibandingkan dengan variabel *event* dan pengujian dengan variabel campuran.
4. Hasil pengujian dataset dua kondisi menggunakan metode *Support Vector Regression (SVR)* mendapatkan nilai yang sangat baik dengan hasil nilai terbaik pada pengujian *splitting 80/20* variabel libur sekolah dengan hasil nilai *MAPE* sebesar 41.3124%. Sedangkan pengujian menggunakan dataset hanya kondisi normal saja tanpa melibatkan dataset ketika *covid* menghasilkan nilai *MAPE* 55.15%. Hasil pengujian ini juga selaras dengan penelitian (Rui Zhang dkk, 2022) tentang jumlah dataset. Dan selaras juga dengan penelitian yang dilakukan oleh (Ahmad Kamal Badri dkk, 2022) tentang pengujian dataset dua kondisi.

5.2. Saran

Hasil penelitian yang telah dilakukan ini disadari oleh penulis masih terlalu jauh dari kata sempurna, diperlukan adanya *improvisasi* lebih lanjut sehingga menghasilkan nilai kinerja yang maksimal dalam melakukan prediksi terhadap

wisatawan. Adapun beberapa saran dari penelitian ini yang dapat diambil adalah sebagai berikut :

1. Untuk penelitian berikutnya supaya mendapatkan nilai kinerja yang lebih maksimal dalam penggunaan metode LSTM diharapkan dataset yang digunakan lebih diperbanyak lagi supaya mendapatkan tingkat kinerja yang lebih akurat.
2. Pada penelitian selanjutnya diperlukan penggunaan metode *preprocessing* yang lain untuk mendapatkan data yang lebih baik dalam melakukan *optimalisasi* pengujian.
3. Pada penelitian berikutnya untuk menguji variabel *event* dan libur sekolah ini diperlukan metode pengujian yang lain baik metode *linear* ataupun metode *non linear* supaya mendapatkan nilai kinerja yang lebih baik.

DAFTAR PUSTAKA

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- 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.
- 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.
- 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.
- 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.
- 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.
- J.-W. Bi, Y. Liu, and H. Li, "Daily tourism volume forecasting for tourist attractions," *Ann Tour Res*, vol. 83, p. 102923, 2020.
- 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.
- Y. Li and H. Cao, "Prediction for tourism flow based on LSTM neural network," *Procedia Comput Sci*, vol. 129, pp. 277–283, 2018.
- 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.
- 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.

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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 MENCAMPURAN KUALITAS BATUBARA MENERAPKAN PERBANDINGAN KOLABORATIF PADA ALGORITMA REGRESI HINGGA NEURAL NET," *Indonesian Mining Professionals Journal*, vol. 4, no. 2, pp. 111–122, 2022.
- S. Lionetti, D. Pfäffli, M. Pouly, T. von der Brück, and P. Wegelin, "Tourism Forecast with Weather, Event, and Crossindustry Data.," in *ICAART (2)*, 2021, pp. 1097–1104.
- N. A. Jauhariyah, H. Habibullo, 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.
- 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.
- 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.
- L. Surtiningsih, M. T. Furqon, and S. Adinugroho, "Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali Menggunakan Support Vector Regression dengan Algoritma Genetika," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 8, pp. 2578–2586, 2018.
- Z. Tan, J. Zhang, Y. He, Y. Zhang, G. Xiong, and Y. Liu, "Short-term load forecasting based on integration of SVR and stacking," *IEEE Access*, vol. 8, pp. 227719–227728, 2020.
- M. Azzahra, B. D. Setiawan, and P. P. Adikara, "Optimasi Parameter Support Vector Regression Dengan Algoritma Genetika Untuk Prediksi Harga Emas," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 2, no. 1, pp. 273–281, 2018.
- D. P. M. Abellana, D. M. C. Rivero, M. E. Aparente, and A. Rivero, "Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: a Philippine scenario," *Journal of Tourism Futures*, vol. 7, no. 1, pp. 78–97, 2021.
- G.-F. Fan, X.-R. Jin, and W.-C. Hong, "Application of COEMD-S-SVR model in tourism demand forecasting and economic behavior analysis: The case of Sanya City," *Journal of the Operational Research Society*, vol. 73, no. 7, pp. 1474–1486, 2022.
- C. Li, P. Ge, Z. Liu, and W. Zheng, "Forecasting tourist arrivals using denoising and potential factors," *Ann Tour Res*, vol. 83, p. 102943, 2020.
- L. Yao, R. Ma, and H. Wang, "Baidu index-based forecast of daily tourist arrivals through rescaled range analysis, support vector regression, and autoregressive integrated moving average," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 365–372, 2021.
- R. Difitria, I. Cholissodin, and I. Indriati, "Penerapan Support Vector Regression dan Particle Swarm Optimization untuk Prediksi Jumlah Kunjungan Wisatawan Mancanegara ke Daerah Istimewa Yogyakarta," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 5, pp. 1364–1371, 2020.

- D. Saputra and M. Ary, "Prediksi Minat Klien Pada Produk Deposito Menggunakan Algoritma SVM Parameter Kernel Polynomial," *eProsiding Sistem Informasi (POTENSI)*, vol. 1, no. 1, pp. 162–171, 2020.
- R. A. Maula et al., "Handling Missing Value dengan Pendekatan Regresi pada Dataset Akuakultur Berukuran Kecil," *Jurnal Rekayasa Elektrika*, vol. 18, no. 3, 2022.
- 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.
- 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.
- 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.
- J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
- Y. Yang, B. Pan, and H. Song, "Predicting hotel demand using destination marketing organization's web traffic data," *J Travel Res*, vol. 53, no. 4, pp. 433–447, 2014.
- Zhang, R., Song, H., Chen, Q., Wang, Y., Wang, S., & Li, Y. (2022). Comparison of ARIMA and LSTM for prediction of hemorrhagic fever at different time scales in China. *Plos one*, 17(1), e0262009.
- Abdoli, G., MehrAra, M., & Ardalani, M. E. Forecasting IRAN's max daily demand for electricity in different weather types using deep learning approach.
- Bathla, G. (2020, November). Stock Price prediction using LSTM and SVR. In 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 211–214). IEEE.
- Khan, S., Ali, M., & Rehman, M. (2020). A Comparative Study of LSTM and Linear Regression for Stock Price Prediction with Noise. *International Journal of Machine Learning and Artificial Intelligence*, 6(2), 22–31.

PUSTAKA LAPORAN PENELITIAN

Kurniawan, W., 2021, Analisis Komparasi Algoritma Support Vector Machine dan Random Forest Untuk Klasifikasi Tingkat Pelanggaran Mahasiswa di Universitas, Tesis, S2 Teknik Informatika, STMIK AMIKOM, Yogyakarta

Maryana, T., 2019, Analisis Time Series Permintaan dan Prediksi Stok dan Nilai Investasi Obat Dengan Menggunakan Metode ABC Analysis dan Support Vector Regression (SVR) Pada Aplikasi "MORBIS", Tesis, S2 Teknik Informatika, STMIK AMIKOM, Yogyakarta

