

SKRIPSI

MULTIVARIATE MULTISTEP TIME SERIES FORECASTING

DENGAN ALGORITMA CNN-LSTM

(Studi Kasus: Data Cuaca di Stasiun BMKG)



Yogi Anggara

NIM. 19106010011

**STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2023

MULTIVARIATE MULTISTEP TIME SERIES FORECASTING
DENGAN ALGORITMA CNN-LSTM
(Studi Kasus: Data Cuaca di Stasiun BMKG)

Skripsi

Untuk memenuhi sebagian persyaratan

Mencapai derajat Sarjana S-1

Program Studi Matematika



Yogi Anggara

NIM. 19106010011

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2023

HALAMAN PERSETUJUAN

Universitas Islam Negeri Sunan Kalijaga

FM-UINSK-BM-05-03/R0

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir
Lamp :

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Yogi Anggara
NIM : 19106010011

Judul Skripsi : *Multivariate Multistep Time Series Forecasting* dengan Algoritma
CNN-LSTM (Studi Kasus: Data Cuaca di Stasiun BMKG)

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 15 Mei 2023

Pembimbing



Dr. Epha Dian Supandi, S.Si., M.Sc.

NIP. 19750912 200801 2 015

HALAMAN PENGESAHAN



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1389/Un.02/DST/PP.00.9/06/2023

Tugas Akhir dengan judul : MULTIVARIATE MULTISTEP TIME SERIES FORECASTING DENGAN ALGORITMA CNN-LSTM (STUDI KASUS: DATA CUACA DI STASIUN BMKG)

yang dipersiapkan dan disusun oleh:

Nama : YOGI ANGGARA
Nomor Induk Mahasiswa : 19106010011
Telah diujikan pada : Jumat, 19 Mei 2023
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Dr. Epha Diana Supandi, S.Si., M.Sc.
SIGNED

Valid ID: 647ff65749aee



Penguji I

Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED

Valid ID: 647f65deb369b



Penguji II

Sri Utami Zuliana, S.Si., M.Sc., Ph.D.
SIGNED

Valid ID: 647ee0126b986



Yogyakarta, 19 Mei 2023
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 648003b6d42dc

HALAMAN PERNYATAAN

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Yogi Anggara
NIM : 19106010011
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 15 Mei 2023


Yogi Anggara



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Karya sederhana ini penulis persembahkan kepada:

Orang tua dan adik penulis yang senantiasa memberikan doa, kasih sayang, dan cinta yang tak terhitung banyaknya.

Keluarga besar, sahabat, dan teman-teman yang selalu memberikan dukungan, inspirasi, dan keceriaan di sepanjang perjalanan hidup ini.

Universitas Islam Negeri Sunan Kalijaga Yogyakarta yang telah memberikan pengalaman dan warna dalam hidup penulis selama empat tahun di program studi Matematika angkatan 2019.



MOTTO

“Tuhan (yang memelihara) dua timur dan Tuhan (yang memelihara) dua barat. Maka nikmat Tuhanmu yang manakah yang kamu dustakan?”

(QS. Ar-Rahman: 17-18)

”Jika Kamu tidak sanggup menahan lelahnya belajar maka kamu harus sanggup menahan perihnya kebodohan.”

(Imam Syafi’i)

”Tak ada penyakit yang tak bisa disembuhkan kecuali kemalasan. Tak ada obat yang tak berguna selain kurangnya pengetahuan.”

(Ibnu Sina)

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PRAKATA



Alhamdulillah, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat, ridha, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "*Multivariate Multistep Time Series Forecasting dengan Algoritma CNN-LSTM (Studi Kasus: Data Cuaca di Stasiun BMKG)*" ini. Shalawat dan salam senantiasa tercurahkan kepada junjungan Nabi Muhammad SAW, yang telah membimbing kita semua menuju jalan yang lurus dan benar.

Penulisan skripsi ini merupakan bagian dari persyaratan akademik yang harus dipenuhi oleh penulis guna memperoleh gelar sarjana dari Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Selama proses penulisan skripsi ini, penulis menyadari bahwa semua ini tidak lepas dari dukungan, bimbingan, dan motivasi dari berbagai pihak. Untuk itu, pada kesempatan ini penulis ingin menyampaikan rasa terimakasih kepada:

1. Bapak Prof. Dr. Phil Al Makin, MA., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
3. Bapak Muchammad Abrori, S.Si., M.Kom., selaku Ketua Program Studi Matematika.
4. Bapak Dr. Sugiyanto, S.Si., S.T., M.Si., selaku Sekretaris Program Studi Matematika sekaligus dosen pembimbing akademik yang telah banyak memberikan pengarahan kepada penulis dari awal hingga akhir masa perkuliahan.
5. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc., selaku dosen pembimbing skripsi yang telah memberikan arahan dan panduan yang sangat berarti bagi penulis dalam menyelesaikan tugas akhir ini serta memperkenalkan penulis dalam dunia kepenulisan.

6. Seluruh dosen program studi Matematika dan staf Fakultas Sains dan Teknologi yang senantiasa memberikan ilmu dan layanan terbaik kepada penulis dari awal hingga akhir perkuliahan. Terkhusus kepada Bapak Arif Munandar, M.Sc. yang telah membimbing penulis dalam olimpiade matematika pada pergelaran OASE 2021 sehingga penulis mampu menorehkan prestasi bagi UIN Sunan Kalijaga sebelum masa kelulusan. Serta kepada Ibu Sri Istiyarti Uswatun Chasanah, M.Si. yang tetap mengizinkan penulis mengikuti perkuliahan Pengantar Matematika Aktuaria meskipun jam perkuliahannya harus terpotong dengan pekerjaan penulis.
7. H. Anda dan Dawiyah Suryaningsih yang merupakan orang tua penulis, serta Zidan Aditya yang merupakan adik penulis, yang telah memberikan dukungan, semangat, dan cinta sehingga menjadi pendorong bagi penulis dalam menyelesaikan skripsi dan mewujudkan cita-cita penulis.
8. Keluarga matematika angkatan 2019 yang telah berjuang dan bertahan bersama selama 4 tahun dalam menyelesaikan pendidikan ini. Kalian semua adalah saudara-saudara yang telah memberikan dukungan, semangat, dan pengalaman berharga selama perjalanan studi kami di bangku kuliah.
9. Bagus, Lathifah, Rosi, Ulil, Hanifah, Putri, dan Rifda selaku teman satu bimbingan penulis yang turut memberikan dukungan dan bantuan selama proses pengerjaan skripsi ini.
10. Teman-teman kelompok KKN Kediri yang selalu suportif selama berlangsungnya KKN sehingga pengabdian tri darma perguruan tinggi dapat berjalan dengan lancar. Terima kasih juga karena meskipun KKN telah berakhir, namun bentuk dukungan teman-teman tetap berlanjut hingga sekarang.
11. Bagus, Dwi, dan Atika sebagai teman dekat penulis yang selalu hadir dalam setiap kesulitan dan kegembiraan selama perjalanan menempuh kuliah dan menyusun skripsi ini. Semangat, motivasi, dan ide-ide segar kalian menjadi bahan bakar utama dalam menyelesaikan skripsi ini.

12. Ibnu, Baskoro, dan Fauzan yang telah memperkenalkan tentang dunia pemrograman dan menjadi teman diskusi penulis tentang teknologi. Tentunya topik skripsi yang penulis ambil tidak lepas dari peran mereka bertiga.
13. Para *programmer open source* yang telah membantu penulis dalam menyelesaikan skripsi ini. Tanpa kontribusinya, mungkin akan membutuhkan waktu lama bagi penulis untuk menyelesaikan skripsi ini.
14. Semua pihak yang tidak bisa penulis sebutkan satu per satu. Semoga selalu dalam lindungan Allah SWT.

Penulis menyadari bahwa dalam penulisan tugas akhir ini, masih terdapat banyak kekurangan dan kelemahan. Oleh karena itu, penulis sangat mengharapkan kritik, saran, dan masukan yang bersifat membangun guna memperbaiki kualitas tugas akhir ini di masa yang akan datang.

Akhir kata, penulis berharap semoga tugas akhir ini dapat bermanfaat bagi semua pihak yang membutuhkan, terutama dalam bidang studi yang terkait. Semoga Allah SWT senantiasa memberikan rahmat, hidayah, dan keberkahan-Nya kepada kita semua. *Aamiin ya Rabbal 'Alamin.*

Yogyakarta, 14 Mei 2023

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA
Yogi Anggara

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN PERSEMBAHAN	v
MOTTO	vi
PRAKATA	vii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR SIMBOL	xvi
DAFTAR LAMPIRAN.....	xviii
INTISARI	xix
ABSTRACT	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Batasan Masalah.....	8
1.3 Rumusan Masalah.....	9
1.4 Tujuan Penelitian.....	9
1.5 Manfaat Penelitian.....	10
1.6 Keaslian Penelitian	10
1.7 Tinjauan Pustaka	10
1.8 Sistematika Penelitian.....	14
BAB II LANDASAN TEORI	16
2.1 Stasiun BMKG Sleman.....	16
2.2 Cuaca	17
2.3 Statistika.....	18

2.4	Analisis Runtun Waktu	20
2.5	Vektor	24
2.6	Matriks	25
2.6.1	Perkalian Matrik	26
2.6.2	<i>Hadamard Product</i>	26
2.6.3	Penggabungan Matriks.....	27
2.6.4	Vektorisasi Matriks.....	27
2.7	Turunan	28
2.7.1	Turunan Rantai.....	29
2.7.2	Turunan Parsial.....	29
2.8	<i>Min-Max Scaler</i>	29
2.9	<i>LOWESS</i>	30
2.10	<i>Machine Learning</i>	31
2.11	<i>Neural Network</i>	36
2.11.1	Perseptron.....	36
2.11.2	Fungsi Aktivasi.....	37
2.11.3	Lapisan <i>Neural Network</i>	39
2.11.4	Formulasi pada <i>Neural Network</i>	40
2.11.5	Fungsi Biaya.....	40
2.11.6	Pelatihan pada <i>Neural Network</i>	41
2.12	CNN-LSTM	44
2.12.1	CNN	44
2.12.2	LSTM	47
2.13	<i>Hyperband</i>	48
2.14	<i>SHapley Additive exPlanation (SHAP)</i>	50
2.15	Metrik Evaluasi	52
BAB III METODOLOGI PENELITIAN.....		54
3.1	Studi Pustaka	54
3.2	Jenis Penelitian	54

3.3	Objek, Variabel, Jenis, dan Sumber Data Penelitian	54
3.4	Metodologi Penelitian	55
3.5	Alat Pengolahan Data	55
3.6	Tahapan Penelitian	55
BAB IV HASIL DAN PEMBAHASAN		59
4.1	Pembentukan Arsitektur CNN-LSTM untuk Model <i>Multivariate Multistep Time Series Forecasting</i>	59
4.2	Nilai Gradien pada Tiap Bobot Lapisan	62
4.3	Algoritma Pelatihan Model CNN-LSTM untuk <i>Model Multivariate Multistep Time Series Forecasting</i>	70
4.4	Pemilihan <i>Hyperparameter</i> Terbaik	70
4.5	Interpretasi Model.....	71
BAB V STUDI KASUS.....		72
5.1	<i>Data Preprocessing</i> dan <i>Feature Engineering</i>	72
5.1.1	Mengatasi Data Kosong	73
5.1.2	Ekstraksi Kecepatan dan Arah Angin Maksimum	73
5.1.3	Enkoding Tanggal Pengamatan.....	75
5.1.4	<i>Smoothing Data</i>	76
5.1.5	Pecah Data Menjadi Data Latih, Validasi, dan Uji.....	78
5.1.6	Normalisasi Data	79
5.1.7	Memecah Data ke Dalam Bentuk Fitur dan Target.....	80
5.2	Penentuan <i>Hyperparameter</i> Terbaik.....	86
5.3	Pelatihan dan Evaluasi Model	88
5.4	Interpretasi Model.....	90
5.5	Analisis Kelebihan dan Kekurangan CNN-LSTM Sebagai Model Peramalan.....	94
BAB VI PENUTUP		96
4.1	Kesimpulan	96
4.2	Saran	96

DAFTAR PUSTAKA	98
LAMPIRAN	103



DAFTAR TABEL

Tabel 2.1 Perbedaan model pada statistika dan machine learning	33
Tabel 4.1 Keterangan data fitur dan data target	60
Tabel 5.1 Data cuaca Stasiun BMKG Sleman	72
Tabel 5.2 Jumlah data kosong	73
Tabel 5.3 Ekstraksi data kecepatan angin maksimum	74
Tabel 5.4 Enkoding tanggal pengamatan	75
Tabel 5.5 Hasil smoothing dengan LOWESS	77
Tabel 5.6 Data latih, validasi, dan uji	78
Tabel 5.7 Nilai minimum dan maksimum data faktor cuaca	80
Tabel 5.8 Hasil perhitungan normalisasi data dengan min-max scaler	80
Tabel 5.9 Data Fitur dan Target.....	81
Tabel 5.10 Sepuluh hyperparameter dengan MAE data validasi terkecil.....	87
Tabel 5.11 MAPE data uji.....	89

DAFTAR GAMBAR

Gambar 2.1 Irisan bidang pada machine learning	32
Gambar 2.2 Arsitektur perseptron	36
Gambar 2.3 Arsitektur neural network	40
Gambar 2. 4 Ilustrasi perhitungan konvolusi terhadap data satu dimensi.....	46
Gambar 2.5 Arsitektur sel memori LSTM	47
Gambar 4.1 Arsitektur CNN-LSTM.....	60
Gambar 5.1 Nilai enkoding data tanggal.....	76
Gambar 5.2 Hasil smoothing data faktor cuaca.....	77
Gambar 5.3 Nilai MAE data pelatihan dan data validasi untuk setiap epoch.....	88
Gambar 5.4 Hasil peramalan data faktor cuaca.....	90
Gambar 5.5 Bar plot untuk 20 fitur terbaik dengan nilai SHAP terbesar	92
Gambar 5.6 <i>Heatmap</i> Rata-Rata Nilai Mutlak SHAP Tiap Pasangan Faktor Cuaca .	93
Gambar 5.7 Plot nilai SHAP untuk setiap lag pada fitur	94

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR SIMBOL

$[\mathbf{a}_i]_{m \leq i \leq n}$	=	Elemen ke- m sampai n pada vektor \mathbf{a}
$(A)_{ij}$	=	Elemen matriks A pada baris ke- i dan kolom ke- j
$[(A)_{ij}]_{i,j \leq n}$	=	Elemen matriks A pada baris dan kolom kurang dari n
$(A)_{i*}$	=	Elemen matriks A pada baris i
$(A)_{*j}$	=	Elemen matriks A pada kolom j
$[A B]$	=	Penggabungan kolom matriks A dan matriks B
$\begin{bmatrix} A \\ - \\ B \end{bmatrix}$	=	Penggabungan baris matriks A dan matriks B
$\text{vec}(A)$	=	Vektorisasi matriks A
\odot	=	Operator <i>hadamard product</i>
\oslash	=	Operator <i>element-wise division</i>
d	=	Notasi turunan
∂	=	Notasi turunan parsial
z_i^l	=	Nilai <i>output</i> dari neuron ke- i lapisan ke- l
a_i^l	=	Nilai fungsi aktivasi dari <i>output</i> neuron ke- i lapisan ke- l
\mathbf{a}^l	=	Nilai fungsi aktivasi seluruh neuron pada lapisan ke- l
ReLU	=	Fungsi <i>Rectified Linear Unit</i>
σ	=	Fungsi sigmoid
tanh	=	Fungsi tangen hiperbolik
sech	=	Fungsi sekan hiperbolik
Conv	=	Notasi perhitungan konvolusi
ConvLayer	=	Notasi lapisan konvolusi
i_t	=	Nilai <i>input gate</i> data sekuensial ke- t
f_t	=	Nilai <i>forget gate</i> data sekuensial ke- t
o_t	=	Nilai <i>output gate</i> data sekuensial ke- t

\tilde{C}_t	=	Kandidat <i>cell state</i>
C_t	=	<i>Cell state</i>
$LSTM_{seq}$	=	Notasi untuk nilai seluruh state lapisan LSTM
ϕ_i	=	Nilai SHAP untuk fitur ke- i
γ	=	Fungsi yang memetakan data pada nilai encoding
C	=	Fungsi biaya (<i>cost function</i>)



STATE ISLAMIC UNIVERSITY
 SUNAN KALIJAGA
 YOGYAKARTA

DAFTAR LAMPIRAN

Lampiran I. Kode sumber <i>data preprocessing</i> dan <i>feature engineering</i>	103
Lampiran II. Kode sumber hyperparameter tuning	108
Lampiran III. Kode sumber pelatihan dan evaluasi model	112
Lampiran IV. Kode sumber pelatihan dan evaluasi model	115
Lampiran V. Biodata Penulis	118



INTISARI

MULTIVARIATE MULTISTEP TIME SERIES FORECASTING

DENGAN ALGORITMA CNN-LSTM

(Studi Kasus: Data Cuaca di Stasiun BMKG)

Oleh

YOGI ANGGARA

19106010011

Analisis runtun waktu adalah analisis kumpulan data dalam rentang waktu tertentu di masa lalu guna memahami dan meramalkan kondisi di masa depan. Umumnya, model klasik pada runtun waktu didesain untuk meramalkan satu langkah ke depan saja. *Neural network* memiliki kemampuan dalam mempelajari model secara umum, sehingga dapat melakukan peramalan beberapa langkah ke depan untuk beberapa data runtun waktu secara sekaligus. Salah satu algoritma pada *neural network* yang dapat digunakan untuk model *multivariate multistep time series* adalah CNN-LSTM. CNN-LSTM (*Convolutional Neural Network - Long Short Term Memory*) merupakan algoritma *neural network* yang memiliki kemampuan dalam mengekstraksi fitur secara otomatis dan mempelajari pola pada data sekuensial tanpa mengabaikan urutannya. Pada penelitian ini, CNN-LSTM digunakan untuk pembentukan model peramalan faktor cuaca di stasiun BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) Kabupaten Sleman bulan Januari 2016 sampai Mei 2022. Data dibagi menjadi data latih, data validasi, dan data uji. Nilai MAPE data uji seluruh faktor cuaca berada di antara 0 sampai 30 persen yang mana hal ini menunjukkan bahwa peramalan sudah cukup layak. Berdasarkan nilai dari SHAP (*SHapley Additive exPlanations*), didapatkan informasi bahwa faktor cuaca yang paling berpengaruh bagi model dalam proses peramalan adalah kelembaban udara.

Kata Kunci: Runtun waktu, multivariat, *Neural network*, CNN-LSTM, SHAP

ABSTRACT

MULTIVARIATE MULTISTEP TIME SERIES FORECASTING

DENGAN ALGORITMA CNN-LSTM

(Studi Kasus: Data Cuaca di Stasiun BMKG)

By

YOGI ANGGARA

19106010011

Time series analysis is the analysis of a collection of data over a certain period of time in the past to understand and predict future conditions. Generally, classic models on time series are designed to forecast only one step ahead. Neural networks have the ability to learn models in general, so they can make forecasts several steps ahead for several time series data simultaneously. One of the neural network algorithms that can be used for multivariate multistep time series models is CNN-LSTM. CNN-LSTM (Convolutional Neural Network - Long Short Term Memory) is a neural network algorithm that has the ability to extract features automatically and learn patterns in sequential data without ignoring the order. In this study, CNN-LSTM was used to build a weather factor forecasting model at the BMKG station (Meteorology, Climatology and Geophysics) Sleman Regency from January 2016 to May 2022. The data is splited into training data, validation data and test data. The MAPE value of the test data for all weather factors is between 0 and 30 percent, which indicates that the forecast is quite feasible. Based on the SHAP (SHapley Additive exPlanations) value, information is obtained that the most influential weather factor for the model in the forecasting process is air humidity.

Keywords: Time series, multivariate, Neural network, CNN-LSTM, SHAP

BAB I

PENDAHULUAN

1.1 Latar Belakang

Statistika adalah disiplin ilmu yang berkaitan dengan pengumpulan, analisis, interpretasi, penjelasan, dan penyajian data (Upton, 2008). Statistika memiliki banyak aplikasi di berbagai bidang seperti ekonomi, administrasi bisnis, perawatan kesehatan, ilmu politik dan banyak lainnya. Statistika digunakan untuk membuat kesimpulan tentang populasi berdasarkan sampel yang diambil dari populasi itu. Metode pada statistika dapat membuat sebuah generalisasi yang lebih baik tentang populasi dan memprediksi perilaku di masa mendatang.

Berdasarkan aktivitas yang dilakukan, statistika dibagi menjadi dua, yaitu statistika deskriptif dan statistika inferensial (Hadi dkk., 2018). Statistika deskriptif dipelopori oleh Florence Nightengale yang merupakan seorang perawat yang sangat inovatif di abad ke-18. Di awal perkembangannya, statistika deskriptif digunakan untuk membangun pemanfaatan grafik dalam penyajian data. Ia memanfaatkan statistika deskriptif untuk menentukan tingkat mortalitas dari data-data yang telah ia kumpulkan. Kerja kerasnya dalam menentukan tingkat mortalitas mampu menurunkan tingkat kematian di masa itu (Supandi, 2020).

Sedangkan statistika inferensial diprakarsai oleh Karl Pearson di abad ke-19, yang juga menjadi permulaan dari ilmu statistika modern. Pearson memperkenalkan pemanfaatan metode dalam statistika untuk penelitian-penelitian di bidang biologi dan social ekonomi. Pearson merupakan peneliti yang berkontribusi dalam pengembangan metode statistika yang sering kali digunakan peneliti-peneliti dalam penelitian kuantitatif, yaitu koefisien korelasi dan analisis regresi yang ia tuliskan pada papernya dengan judul *Mathematical Contributions to the Theory of Evolution*. Selain itu, tokoh modern lainnya adalah Jerzy Neyman, yang memiliki kontribusi sangat besar pada perkembangan uji hipotesis teori peluang, hingga sampling (Supandi, 2020).

Pada abad ke-20, kebutuhan akan analisis statistika semakin kompleks. Salah satu di antaranya adalah bagaimana cara untuk menganalisis data yang bersifat temporal. Metode-metode statistika di abad sebelumnya hanya mampu menganalisis data tabular yang tidak terikat oleh waktu. Meskipun sejak abad ke-18 analisis runtun waktu sudah dipelajari lebih dulu oleh Leonhard Euler pada system dinamik, namun implementasinya di bidang statistika masih belum terealisasikan. Melihat permasalahan tersebut, para ahli statistika kemudian mulai mengembangkan analisis runtun waktu untuk mengatasi permasalahan-permasalahan tersebut. Beberapa tokoh penting dalam perkembangan analisis runtun waktu adalah George Box dan Gwilym Jenkins, yang memperkenalkan metode Box-Jenkins untuk menganalisis data-data yang bersifat temporal. Metode ini menjadi salah satu metode yang paling populer digunakan dalam analisis runtun waktu, karena memiliki kemampuan untuk memprediksi perkembangan suatu sistem di masa yang akan datang dengan tingkat akurasi yang tinggi.

. Selain metode Box-Jenkins, dalam perkembangan selanjutnya terdapat pula metode-metode baru yang dikembangkan dalam analisis runtun waktu, seperti metode ARIMA (*Autoregressive Integrated Moving Average*), metode VAR (*Vector Autoregressive*), dan metode GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*). Metode-metode ini memiliki kemampuan yang lebih baik dalam menganalisis data-data yang bersifat temporal, sehingga dapat digunakan untuk memprediksi perkembangan suatu sistem di masa yang akan datang dengan tingkat akurasi yang lebih tinggi lagi. Sesuai dengan alasan dikembangkannya analisis runtun waktu, model-model tersebut nantinya dapat digunakan untuk berbagai macam tujuan yang berkenaan dengan data temporal, seperti memprediksi perkembangan ekonomi, mengidentifikasi faktor-faktor yang mempengaruhi kinerja suatu perusahaan, menganalisis perkembangan suatu penyakit dalam waktu, atau peramalan cuaca beserta faktornya.

Dewasa ini, perkembangan yang pesat di bidang teknologi informasi menyebabkan turut meningkatnya percepatan penelitian di bidang *deep learning*. *Deep learning* adalah salah satu cabang dari *machine learning* yang memanfaatkan metode

jaringan syaraf tiruan untuk melakukan pembelajaran secara otomatis dari data. Jaringan syaraf tiruan adalah sebuah sistem yang mirip dengan cara kerja otak manusia, dengan menggunakan banyak neuron-neuron yang saling terhubung untuk memproses informasi. *Deep learning* memungkinkan komputer untuk melakukan pembelajaran secara efisien dan cepat dari data yang besar dan kompleks. *Deep learning* telah menjadi teknologi yang sangat populer dalam beberapa tahun terakhir karena kemampuannya untuk menangani berbagai macam masalah yang sebelumnya sulit dipecahkan oleh metode lain. Misalnya, *deep learning* telah digunakan untuk meningkatkan kemampuan komputer dalam melakukan pengenalan wajah, pengenalan suara, dan pemrosesan bahasa alami. *Deep learning* juga telah digunakan dalam bidang medis untuk membantu dokter dalam mengdiagnosa penyakit dan dalam bidang keamanan untuk mengidentifikasi potensi ancaman.

Deep learning bermula dari penelitian tentang jaringan syaraf tiruan yang dimulai pada tahun 1940-an. Pada awalnya, jaringan syaraf tiruan hanya digunakan untuk menyelesaikan masalah yang sederhana, seperti pengenalan pola. Namun, seiring dengan perkembangan teknologi dan metode pembelajaran, jaringan syaraf tiruan semakin dapat menangani masalah yang lebih kompleks. Pada tahun 1980-an dan 1990-an, beberapa peneliti mulai mengembangkan metode pembelajaran yang disebut *backpropagation* yang memungkinkan jaringan syaraf tiruan untuk belajar secara efisien dari data. Kemudian pada tahun 2006, sebuah tim riset dari Universitas Toronto mengembangkan sebuah metode pembelajaran yang menjadi cikal bakal dari *deep learning*, yang memungkinkan jaringan syaraf tiruan untuk memproses data dengan lebih efisien dan mencapai hasil yang lebih baik dalam berbagai macam aplikasi. Sejak saat itu, *deep learning* telah mengalami perkembangan yang pesat dan menjadi salah satu teknologi terpenting dalam bidang *machine learning*.

Pada awalnya, jaringan syaraf tiruan yang digunakan hanya terdiri dari beberapa lapisan yang sederhana, dan tidak bisa menangani data yang kompleks atau berukuran besar. Namun, seiring dengan berjalannya waktu dan perkembangan teknologi, jaringan neural mulai dikembangkan menjadi lebih kompleks dan dapat menangani

data yang lebih besar dan lebih kompleks. *Deep learning* adalah salah satu cabang dari *machine learning* yang berfokus pada penggunaan jaringan neural yang kompleks untuk mengolah data dan membuat prediksi. Perkembangan arsitektur deep learning dimulai pada tahun 1950-an, ketika para ilmuwan mulai menggunakan jaringan neural untuk menyelesaikan masalah-masalah yang kompleks. Pada tahun 1980-an dan 1990-an, beberapa arsitektur deep learning yang penting muncul, termasuk arsitektur *convolutional neural network* (CNN) yang dikembangkan oleh Yann LeCun dan arsitektur *recurrent neural network* (RNN) yang dikembangkan oleh Sepp Hochreiter dan Jürgen Schmidhuber. Kedua arsitektur ini sangat penting dalam perkembangan deep learning, karena memungkinkan jaringan syaraf tiruan untuk menangani data yang memiliki struktur spasial dan waktu, sehingga dapat digunakan untuk memecahkan masalah-masalah seperti pengenalan gambar dan pengolahan bahasa alami.

Meskipun *deep learning* memiliki kelebihan dalam mengenali pola pada data yang begitu kompleks, namun deep learning memiliki kelemahan yang cukup mendasar, yaitu sulitnya penentuan hyperparameter optimal. Hyperparameter adalah parameter yang tidak dapat dioptimalkan secara langsung melalui proses pembelajaran data, tetapi harus ditentukan secara manual oleh para pembuat model. Adapun yang termasuk ke dalam hyperparameter deep learning di antaranya adalah susunan lapisan, banyaknya neuron pada setiap lapisan, dan fungsi aktivasi. Pada awalnya, para pembuat model harus menentukan *hyperparameter* secara manual dan membandingkan hasilnya untuk menemukan parameter terbaik. Hal ini dapat memakan waktu yang lama dan tidak efisien.

Pada tahun 2012, tim riset dari Universitas Google mengembangkan sebuah metode yang disebut "*neural architecture search*" (NAS), yang memungkinkan neural network untuk menemukan parameter terbaik secara otomatis. NAS menggunakan teknik evolusi komputasi untuk mencari *hyperparameter* yang optimal, yaitu dengan cara mencoba berbagai macam kombinasi parameter secara acak dan memilih kombinasi terbaik berdasarkan hasilnya. Dengan menggunakan NAS, para pembuat

model dapat menemukan hyperparameter terbaik suatu model deep learning secara otomatis dan lebih cepat dibandingkan dengan metode manual. Namun, NAS masih membutuhkan waktu yang lama untuk menemukan hasil terbaik, sehingga algoritma hyperband menjadi lebih populer dalam hyperparameter tuning pada deep learning.

Pada tahun 2016, Lisha Li dari Universitas Berkeley mengembangkan algoritma *hyperband*, yang menggabungkan dua metode populer dalam hyperparameter tuning, yaitu *random search* dan *bayesian optimization*. *Hyperband* menggunakan *random search* untuk menemukan parameter yang paling baik dalam jangka pendek, dan kemudian menggunakan *bayesian optimization* untuk memperbaiki hasil tersebut dalam jangka panjang. *Hyperband* dapat menemukan parameter terbaik dengan lebih cepat dibandingkan metode lain, sehingga menjadi salah satu algoritma terpopuler untuk *hyperparameter tuning* pada deep learning. Algoritma ini juga memiliki kemampuan untuk menemukan parameter terbaik dalam jangka pendek, sehingga dapat digunakan untuk mengevaluasi model dengan cepat.

Permasalahan lain dari *deep learning* adalah sulitnya menginterpretasikan model. *Deep learning* umumnya memiliki lapisan dan parameter yang cukup banyak. Hal ini mengakibatkan sulitnya pembuat model dalam menjelaskan variable independent atau fitur apa saja yang dapat membantu model serta seberapa besar fitur tersebut berkontribusi dalam mengenali pola pada data. Akibatnya, meskipun *deep learning* memberikan hasil akurasi yang tinggi dibandingkan model *machine learning* lainnya, *deep learning* tetap sulit diterima oleh para profesional yang berkecukupan dengan hal yang berisiko tinggi seperti bidang kesehatan ataupun keuangan. Oleh sebab itu, peneliti mulai meneliti metode apa yang dapat digunakan untuk menjelaskan cara kerja model *deep learning* secara implisit.

Pada tahun 2014, tim riset dari Universitas Yale mengembangkan metode *permutation importance*, yang merupakan salah satu metode populer untuk mengukur tingkat kepentingan suatu fitur (*feature importance*) pada *deep learning*. Metode ini mengukur *feature importance* dengan cara mengacak fitur-fitur dalam model dan membandingkan hasilnya dengan hasil asli. Dengan menggunakan metode ini, para

pembuat model dapat menemukan fitur-fitur yang paling penting dalam model dengan lebih cepat dan efisien.

Meskipun algoritma *permutation importance* dinilai mampu menjelaskan *feature importance* model *deep learning*, akan tetapi algoritma ini memiliki beberapa kelemahan, di antaranya adalah tidak dapat mengukur interaksi antar fitur dan tidak dapat mengukur *feature importance* secara relatif. Untuk mengatasi kelemahan-kelemahan tersebut, Scott Lundberg dan Su-In Lee dari Universitas Washington mengembangkan metode SHAP (*SHapley Additive exPlanations*) pada tahun 2017. SHAP mengukur *feature importance* dengan menggunakan teori permainan dari Lloyd Shapley, yang memungkinkan metode ini untuk mengukur interaksi antar fitur dan mengukur *feature importance* secara relatif. Sejak dikembangkan, metode SHAP telah menjadi salah satu metode terpopuler untuk mengukur *feature importance* pada *machine learning*, karena kemampuannya untuk mengukur interaksi antar fitur dan mengukur *feature importance* secara relatif.

Penerapan *deep learning* pada model peramalan runtun waktu, terutama pada peramalan cuaca, dapat membantu meningkatkan akurasi prediksi karena jaringan neural dapat menangkap pola yang lebih rumit dari data. Peramalan cuaca merupakan salah satu contoh aplikasi yang banyak menggunakan *deep learning*. Sebagai sebuah model, *deep learning* memiliki kelebihan tersendiri jika dibandingkan dengan model klasik, di antaranya adalah robust terhadap noise dan model yang dihasilkan tidak terpaku pada model linear (Brownlee, 2018). Selain itu, *deep learning* merupakan model non-parametrik yang mana tidak terikat oleh asumsi paramterik apapun. Oleh sebab itu, *deep learning* dapat digunakan sebagai alternative dalam pembentukan model runtun waktu di samping kemampuannya dalam mengenali pola yang kompleks. Sebab, banyak data yang tidak dapat memenuhi asumsi parametrik pada model runtun waktu klasik. Selain itu, kemampuan *deep learning* dalam mengenali pola non-linear memberikan fleksibilitas dalam meramalkan cuaca. Data cuaca memiliki banyak variabel yang saling berkaitan dan dapat menunjukkan pola yang rumit, sehingga memerlukan teknik *machine learning* yang dapat menangkap pola tersebut dengan

baik. Dengan menggunakan deep learning, model peramalan cuaca dapat memprediksi kondisi cuaca di masa depan dengan akurasi yang lebih tinggi. Selain itu, *deep learning* juga memiliki kemampuan untuk menangani data yang berskala besar dan memiliki variasi yang tinggi, sehingga dapat digunakan untuk memperkirakan perubahan cuaca di berbagai wilayah dengan cepat dan akurat. Dengan demikian, penerapan deep learning pada model peramalan runtun waktu, terutama pada peramalan cuaca, dapat membantu meningkatkan efisiensi dan akurasi dalam memprediksi kondisi cuaca di masa depan.

Penelitian-penelitian yang melibatkan *deep learning* dalam peramalan cuaca telah dilakukan oleh banyak peneliti di berbagai negara. Salah satu contohnya adalah penelitian terkait implementasi *multi layer perceptron* untuk peramalan curah hujan jangka pendek. Hasil penelitian menunjukkan bahwa model peramalan tersebut memberikan hasil yang baik berdasarkan nilai *Root Mean Square Error* (RMSE) dan *Threat Score* (TS) (Zhang dkk., 2018). Penelitian lain di antaranya adalah meramalkan suhu udara menggunakan *single-layer perceptron*. Penelitian tersebut juga menganalisis terkait arsitektur jaringan yang dikembangkan, jumlah siklus dalam proses pembelajaran, serta perubahan *Mean Squared Error* (MSE) (Kajewska-Szkudlarek, 2017). Penelitian lainnya menggunakan *Convolutional Neural Network* (CNN) untuk meramalkan kecepatan angin (Harbola & Coors, 2019).

Beberapa penelitian lainnya mencoba menggabungkan beberapa arsitektur *deep learning*, di antaranya adalah penggabungan antara CNN dengan LSTM. Model CNN-LSTM mencoba mengkombinasikan kemampuan yang dimiliki dari kedua arsitektur, yaitu lapisan konvolusi untuk mengotomatisasi ekstraksi data dan lapisan LSTM untuk mempelajari pola data sekuensial. Salah satu contohnya adalah penelitian terkait implementasi SA-CNN-LSTM (Self Attention-Convolutional Neural Network-Long Short Term Memory) dalam meramalkan kecepatan aliran sungai. Hasil dari penelitian tersebut menunjukkan bahwa CNN-LSTM memiliki hasil peramalan yang lebih baik dibandingkan dengan model-model pembanding seperti LSTM, CNN, ANN, RF, SA-LSTM, dan SA-CNN (Zhou dkk., 2023). Penelitian lainnya memanfaatkan CNN-

LSTM untuk meramalkan IHSG (Indeks Harga Saham gabungan) ketika kondisi sedang terjadi COVID-19 dan sebelum COVID-19. Hasil penelitian tersebut menunjukkan bahwa CNN-LSTM tetap mampu meramalkan IHSG dengan baik meskipun pada sebelum COVID-19 maupun pada saat COVID-19 (Anggara & Supandi, 2021).

Cuaca adalah keadaan udara di atmosfer pada waktu dan tempat tertentu, serta tidak stabil dan berubah-ubah. Penentuan klasifikasi cuaca umumnya dinyatakan dengan mempertimbangkan curah hujan harian, suhu, kelembaban udara, tutupan awan, penguapan, kecepatan angin, dan kondisi lokasi lainnya (Aldrian & Budiman, 2011). Pengamatan dan pemahaman mengenai cuaca memiliki signifikansi yang tidak dapat diabaikan dalam berbagai bidang, termasuk meteorologi, pertanian, transportasi, energi, dan penelitian ilmiah. dalam bidang pertanian, pemahaman tentang cuaca dapat mendukung petani dalam mengatur waktu penanaman, pemeliharaan tanaman, dan panen, serta mempersiapkan upaya mitigasi risiko cuaca, seperti kekeringan atau banjir. Selain itu, dalam bidang meteorologi, pemodelan cuaca dan prakiraan cuaca memegang peranan penting dalam memahami dinamika atmosfer serta menyediakan informasi yang berharga dalam perencanaan dan pengelolaan aktivitas sehari-hari, termasuk pengaturan lalu lintas udara, navigasi maritim, dan kegiatan luar ruangan.

Pada skripsi ini, penulis akan melakukan penelitian terkait implementasi dari CNN-LSTM dalam pembentukan model *multivariate time series forecasting* pada data cuaca stasiun BMKG Kabupaten Sleman. Setelah didapatkan hasil model yang baik berdasarkan evaluasi nilai MSE dan MAPE, berikutnya akan ditentukan *feature importance* model dengan memanfaatkan algoritma SHAP guna memudahkan proses interpretasi.

1.2 Batasan Masalah

Pembatasan masalah perlu dilakukan supaya penelitian ini dapat mendekati sasaran yang diharapkan, diantaranya adalah:

1. Data yang digunakan pada penelitian ini adalah data cuaca tanggal 1 Januari 2016 sampai dengan 31 Mei 2022 dari stasiun BMKG Kabupaten Sleman, Yogyakarta.
2. Pembentukan model *multivariat multistep time series* menggunakan arsitektur CNN-LSTM
3. *Hyperparameter* tuning menggunakan algoritma *hyperband*
4. *Feature importance* menggunakan SHAP
5. Nilai evaluasi menggunakan *mean absolute error* untuk data validasi dan *mean absolute percentage error* untuk data uji
6. Bahasa program yang digunakan pada penelitian ini adalah Python versi 3.8.0 di mana pustaka yang digunakan di antaranya adalah Tensorflow untuk pembentukan model CNN-LSTM, Pandas dan Numpy untuk prapemrosesan data, dan Matplotlib untuk visualisasi data.

1.3 Rumusan Masalah

1. Bagaimana cara melakukan prapemrosesan data runtun waktu untuk model CNN-LSTM pada kasus data cuaca
2. Bagaimana cara mengimplementasikan CNN-LSTM untuk melakukan pemodelan multivariat multistep time series
3. Bagaimana cara menentukan parameter optimal pada CNN-LSTM
4. Bagaimana cara menentukan *feature importance* pada CNN-LSTM
5. Apa kelebihan dan kekurangan CNN-LSTM sebagai model peramalan

1.4 Tujuan Penelitian

Tujuan disusunnya skripsi ini adalah:

1. Untuk memahami pemanfaatan arsitektur CNN-LSTM dalam pembentukan model *multistep multivariate time series forecasting* pada kasus data cuaca
2. Untuk memahami cara penentuan *hyperparameter* optimal menggunakan algoritma *hyperband*

3. Untuk mengetahui fitur apa saja yang penting bagi CNN-LSTM dalam meramalkan data cuaca menggunakan SHAP
4. Untuk menganalisis kelebihan dan kekurangan CNN-LSTM sebagai model peramalan

1.5 Manfaat Penelitian

Kebermanfaatan skripsi ini bagi penulis adalah meningkatkan pemahaman penulis dalam pengembangan deep learning menggunakan CNN-LSTM mulai dari prapemrosesan data runtun waktu hingga interpretasi model menggunakan algoritma SHAP dan permutation importance. Penelitian pada skripsi ini juga bermanfaat bagi penulis sebagai portofolio penulis di bidang sains data. Penelitian ini juga dapat menjadi referensi bagi pembaca yang tengah mempelajari *machine learning*. Selain itu, model yang dibuat pada penelitian ini juga bisa dijadikan sebagai model pendukung dalam peramalan cuaca untuk BMKG maupun instansi terkait.

1.6 Keaslian Penelitian

Penelitian terkait penerapan CNN-LSTM dalam pembentukan model *multivariate multistep time series forecasting* pada data cuaca BMKG Kabupaten Sleman belum pernah dilakukan. Akan tetapi, pemanfaatan CNN-LSTM dalam peramalan pernah diteliti sebelumnya. Yang membedakan dengan penelitian ini adalah studi kasus, banyaknya variabel masukan dan keluaran, serta metode interpretasi modelnya.

1.7 Tinjauan Pustaka

Tinjauan pustaka berisi tentang kajian literatur dari penelitian terdahulu. Beberapa referensi utama di dalam bab ini, yaitu sebagai berikut:

1. Joanna Kajewska-Szkudlarek (Kajewska-Szkudlarek, 2017), melakukan penelitian terkait peramalan suhu udara menggunakan *single-layer perceptron*. Kajian yang dianalisis pada penelitian ini terkait arsitektur *neural network*, banyaknya siklus

selama proses pelatihan, perubahan nilai *Mean Square Error* (MSE), korelasi antara nilai parameter yang didapatkan terhadap instrumen meteorologi, dan analisis sensitivitas global. Pada penelitian ini, banyak neuron pada *hidden layer* yang digunakan pada kedua arsitektur yang digunakan untuk eksperimen adalah sebanyak 4 dan 6 buah. Nilai MSE data uji pada kedua arsitektur tersebut berturut-turut adalah 0,833 dan 0,724. Nilai korelasi data uji antara nilai parameter dengan instrumen meteorologi dari kedua arsitektur berturut-turut adalah 0,995 dan 0,996 yang mana menunjukkan bahwa adanya korelasi positif yang sangat kuat. Analisis sensitivitas global menunjukkan peran paling signifikan dari suhu rata-rata dan kemudian suhu maksimum dan minimum diukur dengan metode standar dalam model yang dibuat.

2. Pengcheng Zhang, Yangyang Jia, Jerry Gao, Wei Song, Hareton Leung (Zhang dkk., 2018), melakukan penelitian terkait peramalan curah hujan di Cina menggunakan *Multi-Layer Perceptron* (MLP). Input yang diberikan pada MLP adalah data reduksi ketiga belas faktor fisik menggunakan *Principal Component Analysis* (PCA). Pelatihan MLP menggunakan algoritma *greedy*. Hasil percobaan dari penelitian tersebut memberikan hasil peramalan yang baik berdasarkan nilai *Root Mean Square Error* (RMSE) dan *Threat Score* (TS).
3. Luqman Affandi, Arie Rachmad Syulistyo, dan Firdani Rianda Putra (Affandi dkk., 2018), melakukan penelitian terkait pengembangan aplikasi *mobile* peramalan cuaca untuk penentuan tanaman pangan yang sesuai dengan kondisi cuaca pada periode ke depannya bagi para petani. Faktor-faktor penentu cuaca, seperti suhu, cuaca, curah hujan, dan lama penyinaran matahari, diprediksi secara terpisah menggunakan metode *triple exponential smoothing*. Prediksi tersebut berikutnya menjadi acuan dalam penentuan rekomendasi tanaman pangan yang cocok terhadap cuaca di beberapa periode ke depan. Hasil penelitian tersebut menunjukkan bahwa nilai *Mean Absolute Deviation* (MAD) yang digunakan sebagai metrik evaluasi memberikan hasil yang cukup baik, di mana nilai MAD terkecil untuk data suhu

adalah sebesar 0,705 untuk $\alpha = 0,1$, nilai MAD terkecil untuk data kelembapan adalah sebesar 4,523 untuk $\alpha = 0,4$, dan nilai MAD terkecil untuk data curah hujan adalah sebesar 129,120 untuk $\alpha = 0,1$.

4. Fahrur Rozi dan Farid Sukmana (Rozi & Sukmana, 2016), melakukan penelitian terkait model peramalan cuaca menggunakan metode siklis dan adaptive neuro fuzzy inference system dari data BMKG Karangploso, Malang. Parameter pengaruh kondisi cuaca yang digunakan pada penelitian ini di antaranya adalah suhu, tekanan udara, kelembapan udara, dan kecepatan angin. Model peramalan cuaca yang dibangun merupakan model klasifikasi, di mana output labelnya adalah cerah, mendung, dan hujan. Pada penelitian ini, langkah pertama yang dilakukan adalah mengelompokkan data menggunakan K-Means Clustering. Berikutnya melakukan metode ANFIS dengan menggunakan metode aturan Sugeno dengan mengimplementasikannya ke dalam bentuk *artificial neural network* (ANN) dengan jumlah lapisan sebanyak 5 buah, yang mana terdiri dari 3 lapisan tetap dan 2 lapisan adaptif. Metode siklis kemudian diterapkan saat membangkitkan nilai parameter cuaca untuk periode waktu $t+1$. Nilai akurasi terbaik yang diperoleh dari model peramalan cuaca pada penelitian ini sebesar 76,67% dengan menggunakan data selama tiga bulan sebagai data latih dan laju pembelajaran sebesar 0,3.
5. Risty Jayanti Yuniar, Didik Rahadi S. dan Onny Setyawati (Yuniar dkk., 2013), melakukan penelitian terkait perbaikan metode prakiraan cuaca dengan algoritma Jaringan Syaraf Tiruan Propagasi Terbalik di Bandara Abdulrahman Saleh. Penelitian ini secara spesifik meramalkan dua faktor penentu cuaca, yaitu curah hujan dan kecepatan angin. Sedangkan parameter yang digunakan sebagai input adalah kelembapan udara, suhu udara, dan tekanan udara. Model terbaik didapatkan saat peneliti menggunakan 5 lapisan tersembunyi (*hidden layer*) dan kecepatan pembelajaran sebesar 0,9. Nilai MSE yang didapatkan pada penelitian ini adalah sebesar 0,005 untuk prakiraan curah hujan dan 0,009 untuk prakiraan kecepatan angin.

6. Dinita Rahmalia dan Teguh Herlambang (Rahmalia & Herlambang, 2017), melakukan penelitian terkait prediksi cuaca dengan memanfaatkan algoritma *Particle Swarm Optimization-Neural Network* (PSONN). Kegunaan algoritma PSO pada penelitian ini adalah untuk menentukan bobot optimal mula-mula jaringan syaraf tiruan. Bobot tersebut kemudian diperbarui dengan memanfaatkan algoritma propagasi maju dan propogasi mundur pada jaringan syaraf tiruan. Hasil simulasi menunjukkan bahwa algoritma PSONN dapat menghasilkan nilai prediksi cuaca yang mendekati nilai target.
7. Shubhi Harbola dan Volker Coors (Harbola & Coors, 2019), melakukan peramalan kecepatan dan arah angin pada daerah Stuttgart, Jerman, menggunakan CNN (*Convolution Neural Network*) satu dimensi. Akurasi yang didapatkan pada penelitian ini sebesar 95,2%. Menurut peneliti, hasil dari penelitian ini akan sangat membantu dalam pemasangan turbin angin yang output dayanya bergantung pada parameter tersebut.
8. Eko Supriyadi (Supriyadi, 2021), melakukan penelitian terkait prediksi parameter cuaca dengan menggunakan *Long-Short Term Memory* (LSTM) dengan lapisan (*layer*) sebanyak 200 buah. Data yang digunakan adalah suhu udara, kelembaban udara, kecepatan angin dan tekanan udara dari bulan Januari sampai Februari 2019, di mana data bulan Januari sebagai data latih dan data uji, sedangkan data bulan Februari digunakan sebagai data pembandingan hasil prediksi. Hasil dari penelitian ini menunjukkan bahwa nilai *Root Mean Square Error* (RMSE) setiap parameter cuaca memberikan hasil yang cukup baik, di mana untuk data suhu diperoleh nilai RMSE sebesar 0,576, untuk data kelembapan diperoleh nilai RMSE sebesar 2,869, unuk data kecepatan angin diperoleh nilai RMSE sebesar 2,196, dan data tekanan udara diperoleh nilai RMSE sebesar 1,431. Peneliti tersebut juga menunjukan bahwa hanya data suhu dan kelembaban udara yang mengalami peningkatan nilai RMSE seiring bertambahnya waktu, sedangkan data lainnya mengalami penurunan nilai RMSE di hari ketiga dan kemudian meningkat kembali hingga satu bulan ke depan.

9. Yogi dan Epha (Anggara & Supandi, 2021), melakukan penelitian terkait permalan Indeks Harga Saham Gabungan (IHSG) dengan memanfaatkan CNN-LSTM. Pada penelitian ini, penentuan neuron optimal di tiap lapisan menggunakan algoritma *hyperband*. Hasil penelitian menunjukkan bahwa CNN-LSTM mampu meramalkan IHSG dengan akurat dalam memprediksi satu hari ke depan, baik sebelum ataupun saat pandemic COVID-19, di mana besarnya MAPE sebelum COVID-19 sebesar 1,4% dan semasa COVID-19 sebesar 0,5%. Pada penelitian tersebut juga diperoleh informasi bahwa fitur yang paling mempengaruhi model dalam meramalkan data adalah data IHSG lag pertama.

Pada skripsi ini penulis meneliti terkait penerapan *Convolutional Neural Network – Long-Short Term Memory* (CNN-LSTM) dalam pembentukan model peramalan multivariate multistep time series pada data cuaca yang dipublikasikan oleh Stasiun BMKG Sleman. Perbedaan dari penelitian sebelumnya adalah model pada penelitian ini mampu memprediksi factor-faktor pembentuk cuaca beberapa periode ke depan sekaligus secara serentak. Faktor-faktor pembentuk cuaca yang diteliti pada skripsi ini di antaranya adalah suhu rata-rata, kelembaban rata-rata, curah hujan, intensitas cahaya, dan kecepatan angin rata-rata. Pada penelitian ini, fungsi optimasi yang digunakan adalah Adam, yang mana merupakan algoritma untuk optimasi fungsi objektif stokastik yang menggunakan gradien orde pertama sebagai basisnya. Metrik evaluasi yang digunakan adalah *Mean Absolute Error* (MAE) untuk data validasi dan *Mean Absolute Percentage Error* (MAPE) untuk data uji. Karena pada penelitian ini semua variable dimodelkan secara serentak, maka *feature importance* antar variable yang mempengaruhi prediksi model dapat diukur. Adapun algoritma yang digunakan untuk mengukur *feature importance* pada penelitian ini adalah SHAP (*Shapley Additive Explanations*).

1.8 Sistematika Penelitian

BAB I PENDAHULUAN

Dalam bab ini berisi latar belakang, tujuan penelitian, rumusan masalah, batasan masalah, manfaat penelitian, dan sistematika penelitian.

BAB II LANDASAN TEORI

Dalam bab ini menjelaskan tentang konsep-konsep teori yang menjadi acuan, berisi tentang landasan teori, hipotesis, dan rancangan penelitian.

BAB III METODE PENELITIAN

Dalam bab ini membahas tentang metode penelitian yang digunakan dalam memecahkan masalah, di mana berisi tentang pendekatan penelitian, sumber data dan variabel penelitian.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini akan dilakukan analisis data sampai ditemukan hasilnya yang kemudian dibahas secara terperinci.

BAB V STUDI KASUS

Dalam bab ini membahas analisis terkait studi kasus dengan memanfaatkan metode yang telah dipaparkan pada BAB IV.

BAB VI PENUTUP

Dalam bab ini berisi kesimpulan dari penelitian yang dilakukan kemudian ditambahkan saran dari penulis.

BAB VI

PENUTUP

4.1 Kesimpulan

1. Algoritma CNN-LSTM adalah algoritma pada *neural network* yang menggabungkan lapisan konvolusi dengan LSTM sehingga memiliki kemampuan dalam mengekstraksi fitur dan mempelajari pola non-linear pada data sekuensial
2. Pelatihan model CNN-LSTM dengan menggunakan konfigurasi hyperparameter yang didapatkan dari algoritma hyperband memiliki nilai MAE data latih dan data validasi yang kecil dan terus mengalami penurunan pada setiap kenaikan epoch. Dengan demikian, hyperband merupakan salah satu algoritma yang cocok digunakan untuk *neural network* dengan arsitektur CNN-LSTM karena mampu menghindari model dari terjadinya *overfitting dan underfitting*.
3. Banyak neuron pada LSTM lebih banyak dibandingkan dengan banyaknya filter konvolusi. Hal ini mengindikasikan bahwa model membutuhkan banyak neuron untuk mengenali pola sekuensial pada data cuaca. Dengan demikian, dapat dikatakan bahwa pola sekuensial pada data cuaca sangat penting bagi model dalam peramalan.
4. Kelembaban rata-rata memberikan kontribusi paling besar bagi model dalam meramalkan faktor cuaca lainnya. Sedangkan suhu maksimum dan lama penyinaran matahari memberikan kontribusi paling kecil bagi model dalam meramalkan faktor cuaca lainnya.

4.2 Saran

1. Hyperparameter yang dioptimasi pada penelitian ini terbatas hanya pada banyaknya neuron pada tiap layer CNN-LSTM dan besarnya *learning rate*. Pada penelitian berikutnya sebaiknya dapat menambahkan beberapa konfigurasi hyperparameter yang akan dioptimasi sehingga kemampuan model dalam meramalkan menjadi lebih baik lagi.

2. Penelitian ini hanya menggunakan arsitektur CNN-LSTM yang sederhana. Pada penelitian berikutnya dapat dieksplorasi lebih lanjut lagi terkait bentuk arsitektur CNN-LSTM yang jauh lebih baik lagi dalam menangani data cuaca.
3. Penelitian ini hanya menggunakan data 7 hari pertama sebagai data input untuk meramalkan 2 hari ke depan. Pada penelitian berikutnya dapat dieksplorasi lebih jauh lagi terkait berapa banyak lag yang dibutuhkan dan banyaknya langkah yang diramalkan CNN-LSTM agar tetap mampu meramalkan data cuaca dengan baik.
4. Pada penelitian ini peramalan hanya terbatas di BMKG Bantul saja. Pada penelitian berikutnya dapat ditambahkan fitur koordinat sehingga model dapat meramalkan tidak hanya pada satu daerah saja.



DAFTAR PUSTAKA

- Affandi, L., Syulistyo, A. R., & Putra, F. R. (2018). Pengembangan Aplikasi Mobile Peramalan Cuaca Untuk Penentuan Tanaman Pangan Menggunakan Metode Triple Exponential Smoothing. *Jurnal Informatika Polinema*, 4(2), 117. <https://doi.org/10.33795/jip.v4i2.155>
- Aldrian, E., & Budiman, M. (2011). *Adaptasi dan mitigasi perubahan iklim di Indonesia*. November 2011. https://www.researchgate.net/profile/Edvin_Aldrian/publication/309721670_Adaptasi_dan_Mitigasi_Perubahan_Iklim_di_Indonesia/links/581ec39c08aea429b295db6b.pdf
- Ananda, R., & Fadhli, M. (2018). *Statistik Pendidikan: Teori dan Praktik dalam Pendidikan*. Widya Puspita.
- Anggara, Y., & Supandi, E. D. (2021). Jakarta Composite Index Model Before and During COVID-19 Using CNN-LSTM. *Proceedings of the International Conference on Science and Engineering (ICSE-UIN-SUKA 2021)*, 211, 226–232. <https://doi.org/10.2991/AER.K.211222.036>
- Anton, H., & Rorres, C. (2005). *Elementary Linear Algebra* (9 ed.). Wiley & Sons.
- Ayo, C. K. (2014). *Stock Price Prediction Using the ARIMA Model*. <https://doi.org/10.1109/UKSim.2014.67>
- Badan Meteorologi, K. dan G. (t.t.). *Pusat Database BMKG*. Diambil 12 Mei 2023, dari <https://dataonline.bmkg.go.id/home>
- Barber, D. (2012). Statistics for machine learning. Dalam *Bayesian Reasoning and Machine Learning*. <https://doi.org/10.1017/cbo9780511804779.012>
- Berger, J. A., Hautaniemi, S., Järvinen, A. K., Edgren, H., Mitra, S. K., & Astola, J. (2004). Optimized LOWESS normalization parameter selection for DNA microarray data. *BMC Bioinformatics*, 5(1), 1–13. <https://doi.org/10.1186/1471-2105-5-194/FIGURES/5>

- Chatfield, C. (1984). *The Analysis of Time Series: An Introduction* (3 ed., Vol. 4, Nomor 1). Chapman and Hall Ltd.
- Christopher M. Bishop. (2016). Pattern Recognition and Machine Learning. Dalam *Springer*. https://doi.org/10.1007/978-3-030-57077-4_11
- Dwitanto, D. S. (2011). *Analisis Runtun Waktu untuk Meramalkan Jumlah Pasien yang Berobat di Puskesmas Blora dengan Menggunakan Software Minitab 14*. Universitas Negeri Semarang.
- Gentle, J. E. (2007). *Matrix Algebra: Theory, Computations, and Applications in Statistics*.
- Hadi, S., Gunawan, I., & Dalle, J. (2018). *Statistika Inferensial: Teori dan Aplikasinya* (2 ed.). Rajagrafindo Persada.
- Harbola, S., & Coors, V. (2019). One Dimensional Convolutional Neural Network Architectures for Wind Prediction. *Energy Conversion and Management*, 195, 70–75. <https://doi.org/10.1016/j.enconman.2019.05.007>
- Hastie, T., Tibshirani, R., & Friedman, J. (2008). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Dalam *Revista Espanola de las Enfermedades del Aparato Digestivo* (2 ed.). Springer.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Kajewska-Szkudlarek, J. (2017). Neural Network Modeling of Automatic Air Temperature Time Series. *Italian Journal of Agrometeorology*, 22(2), 5–12. <https://doi.org/10.19199/2017.2.2038-5625.005>
- Kanal, F. A., Manurung, T., & Prang, J. D. (2018). Penerapan Model Garch (Generalized Autoregressive Conditional Heteroscedasticity) Dalam Menghitung Nilai Beta Saham Indeks Pefindo25. *Jurnal Ilmiah Sains*, 18(2), 67. <https://doi.org/10.35799/jis.18.2.2018.19732>
- Lecun, Y., Bottou, L., Bengio, Y., & Ha, P. (1998). GradientBased Learning Applied to Document Recognition. *Proceedings of the IEEE*, November, 1–46.

- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2018a). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, *18*, 1–52.
- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2018b). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, *18*, 1–52.
- Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. *Complexity*, *2020*. <https://doi.org/10.1155/2020/6622927>
- Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *2017-December*, 4766–4775. <https://doi.org/10.48550/arxiv.1705.07874>
- Musthofa, M. U., Umma, Z. K., & Handayani, A. N. (2017). Analisis Jaringan Saraf Tiruan Model Perceptron Pada Pengenalan Pola Pulau di Indonesia. *Jurnal Ilmiah Teknologi Informasi Asia*, *11*(1), 89. <https://doi.org/10.32815/jitika.v11i1.56>
- Nasution, D., Harumy, T. H. F., Haryanto, E., Fachrizal, F., Julham, & Turnip, A. (2016). A classification method for prediction of qualitative properties of multivariate EEG-P300 signals. *Proceedings of the 2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology, ICACOMIT 2015*, 82–86. <https://doi.org/10.1109/ICACOMIT.2015.7440180>
- Nawawi, G. (2001). Pengantar Klimatologi Pertanian. *Departement Pendidikan Nasional*, 1–44.
- Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press.
- Observations made at secondary stations in the Netherlands Indies*. (1913). <https://catalogue.nla.gov.au/Record/2047221>
- Peraturan Kepala Badan Meteorologi, Klimatologi, dan Geofisika*. (2014). https://iklim.bmkg.go.id/publikasi-klimat/ftp/regulasi-brosur/Perka_10.2014_UT_Staklim.pdf

- Perkiraan deret waktu* / *TensorFlow Core*. (t.t.). Diambil 7 Februari 2023, dari https://www.tensorflow.org/tutorials/structured_data/time_series?hl=id#next_steps
- Pujiastuti, C. E., & Riyono, J. (2022). Pelatihan Pembuatan Analisis Runtun Waktu untuk Optimasi Produk dengan Minitab. *JURNAL ABDI MASYARAKAT INDONESIA (JAMIN)*, 129–136.
- Rahmalia, D., & Herlambang, T. (2017). Prediksi Cuaca Menggunakan Algoritma Particle Swarm Optimization-Neural Network (PSO-NN). *Seminar Nasional Matematika dan Aplikasinya (SNMA)*, 41–48.
- Retno, D., Saputro, S., Wigena, A. H., & Djuraidah, A. (2011). Model Vektor Autoregressive Untuk Peramalan Curah Hujan di Indramayu. *Forum Statistika dan Komputasi*, 16(2), 7–11.
- Rozi, F., & Sukmana, F. (2016). Metode Siklis Dan Adaptive Neuro Fuzzy Inference System Untuk Peramalan Cuaca. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 1(01), 7–13. <https://doi.org/10.29100/jipi.v1i01.20>
- Rudin, C., & Radin, J. (2019). Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition. *Harvard Data Science Review*, 1(2), 2019. <https://doi.org/10.1162/99608F92.5A8A3A3D>
- Shapley, L. S. (1953). A Value for n-Person Games. *Contributions to the Theory of Games (AM-28)*, Volume II, 307–318. <https://doi.org/10.1515/9781400881970-018/HTML>
- Soejoeti, Z. (1987). *Analisis Runtun Waktu*. Karunika Universitas Terbuka.
- Staklim Yogyakarta – Website Resmi Stasiun Klimatologi Yogyakarta*. (t.t.). Diambil 17 Mei 2023, dari <https://staklimyogyakarta.com/>
- Sumarjaya, I. W. (2016). Modul Analisis Deret Waktu. Dalam *Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Udayana*. Universitas Udayana Press.
- Supandi, E. D. (2020). *Statistika dan Terapannya*. Refika.

- Supriyadi, E. (2021). Prediksi Parameter Cuaca Menggunakan Deep Learning Long-Short Term Memory (Lstm). *Jurnal Meteorologi dan Geofisika*, 21(2), 55. <https://doi.org/10.31172/jmg.v21i2.619>
- Suryani, U., & Wakhid, M. (2018). Permainan Kooperatif Bentuk Koalisi. *Jurnal Logika*, 2, 80–86.
- Suyanto. (2014). *Artificial Intelligence: Searching, Reasoning, Planning and Learning*. Informatika.
- Ta, V.-D., Liu, C.-M., & Tadesse, D. A. (2020). Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. *Applied Sciences* 2020, Vol. 10, Page 437, 10(2), 437. <https://doi.org/10.3390/APP10020437>
- Upton, G. (2008). *A Dictionary of Statistics* (2 ed.). Oxford University Press.
- Wirawan, N. (2016). *Statistika Ekonomi dan Bisnis: Statistika Deskriptif. Dalam Statistika Deskriptif*. Keraras Emas.
- Yanto, M., Sovia, R., & Wiyata, P. (2018). Jaringan Syaraf Tiruan Perceptron untuk Penentuan Pola Sistem Irigasi Lahan Pertanian di Kabupaten Pesisir Selatan Sumatra Barat. *Sebatik*, 111–115.
- Yuniar, R. J., S., D. R., & Setyawati, O. (2013). Perbaikan Metode Prakiraan Cuaca Bandara Abdulrahman Saleh Dengan Algoritma Neural Network Backpropagation. *Jurnal EECCIS*, 7(1), pp.65-70.
- Zhang, P., Jia, Y., Gao, J., Song, W., & Leung, H. (2018). Short-Term Rainfall Forecasting Using Multi-Layer Perceptron. *IEEE Transactions on Big Data*, 6(1), 93–106. <https://doi.org/10.1109/tbdata.2018.2871151>
- Zhou, F., Chen, Y., & Liu, J. (2023). Application of a New Hybrid Deep Learning Model That Considers Temporal and Feature Dependencies in Rainfall–Runoff Simulation. *Remote Sensing*, 15(5). <https://doi.org/10.3390/rs15051395>