

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**PREDIKSI KUALITAS UDARA DI KOTA PEKANBARU
MENGUNAKAN ALGORITMA *LONG SHORT TERM
MEMORY* (LSTM)**

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

REZKI QURATUL NF

NIM. 12050110320



UIN SUSKA RIAU

UIN SUSKA RIAU

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2024**

LEMBAR PERSETUJUAN

PREDIKSI KUALITAS UDARA DI KOTA PEKANBARU MENGUNAKAN ALGORITMA *LONG SHORT TERM MEMORY* (LSTM)

TUGAS AKHIR

Oleh

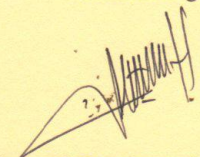
REZKI QURATUL NF

NIM. 12050110320

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 14 November 2024

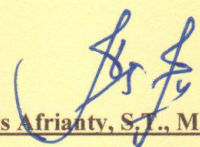
Pembimbing I,

Pembimbing II



Dr. Lestari Handayani, S.T., M. Kom

NIP. 19811113 200710 2 003



Iis Afrianty, S.T., M. Sc

NIP. 19880426 201903 2 009

LEMBAR PENGESAHAN

PREDIKSI KUALITAS UDARA DI KOTA PEKANBARU MENGUNAKAN ALGORITMA *LONG SHORT TERM MEMORY* (LSTM)

Oleh

REZKI QURATUL NF

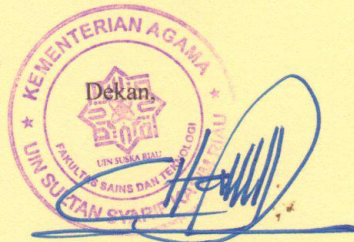
NIM. 12050110320

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 14 November 2024

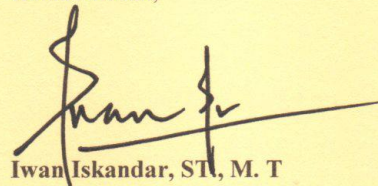
Mengesahkan,

Ketua Jurusan,



Dr. Hartono, M. Pd

NIP. 19640301 199203 1 003



Iwan Iskandar, ST, M. T

NIP. 19821216 201503 1 003

DEWAN PENGUJI

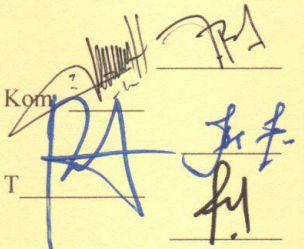
Ketua : Novriyanto, ST., M. Sc

Pembimbing I : Dr. Lestari Handayani, S.T., M. Kom

Pembimbing II : Iis Afrianty, S. T., M. Sc

Penguji I : Dr. Rahmad Abdillah, S. T., M. T

Penguji II : Lola Oktavia, S. S. T., M. T. I



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 14 November 2024

Yang membuat pernyataan,



REZKI QURATUL NF

12050110320

LEMBAR PERSEMBAHAN

Tulisan pada lembar persembahan merupakan ekspresi bebas mahasiswa yang dimaksudkan sebagai ungkapan filosofis yang dapat memotivasi semua pihak, khususnya mahasiswa/penulis sendiri dalam penyelesaian tugas akhir.

Lembar persembahan format bebas, tidak berlebih-lebihan dan dalam komposisi warna grayscale atau hitam-putih.

Lembar persembahan dibuat secara wajar, tidak menyinggung, tidak mengandung hoaks dan tidak mengandung unsur SARA.

Lembar persembahan cukup dibuat 1 halaman ini saja.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

ABSTRAK

Kualitas udara merupakan indikator tingkat kebersihan atau pencemaran udara oleh berbagai zat polutan. Kota Pekanbaru sebagai salah satu kota besar di Indonesia yang menghadapi permasalahan pencemaran udara, terutama akibat kebakaran hutan dan lahan serta aktivitas transportasi. Hal ini berdasarkan data dari Dinas Lingkungan Hidup kota Pekanbaru pada tahun 2018 hingga 2023, yang mencakup parameter PM_{10} , $PM_{2.5}$, SO_2 , NO_2 , CO , dan O_3 . Penelitian ini bertujuan untuk memprediksi kualitas udara di Kota Pekanbaru menggunakan algoritma *Long Short Term Memory* (LSTM) untuk membangun model prediksi data *time series* kualitas udara. Algoritma ini dipilih karena mampu menghasilkan akurasi tinggi dan error evaluasi yang rendah. Total data yang digunakan 2190 berupa data harian. Skenario pengujian yang dilakukan dengan pembagian data uji dan latih 80:20, dan 90:10, serta penanganan *missing value* yang dilakukan *mean*, *median*, dan *interpolasi linear*, dan jumlah *epoch* 50, 100, dan 150. Evaluasi model dilakukan dengan Metrik *Absolute Percentage Errors* (MAPE) dan *Root Squared Error* (RMSE). Hasil pengujian menunjukkan bahwa skenario terbaik diperoleh dengan metode *interpolasi linear* untuk menangani *missing value*, pembagian data pelatihan dan pengujian dengan rasio 90:10, serta penggunaan 100 *epoch*. Model prediksi yang dihasilkan menunjukkan nilai RMSE dan MAPE yang relatif kecil, menunjukkan tingkat akurasi yang cukup memuaskan.

Kata Kunci: Kualitas udara, Prediksi, *Long Short Term Memory*, Polusi udara, Pekanbaru.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

ABSTRACT

Air quality is an indicator of the level of cleanliness or pollution in the air caused by various pollutants. Pekanbaru, as one of the major cities in Indonesia, faces significant air pollution problems, primarily due to forest and land fires as well as transportation activities. This is based on data from the Pekanbaru City Environmental Service from 2018 to 2023, which includes parameters such as PM_{10} , $PM_{2.5}$, SO_2 , NO_2 , CO , and O_3 . This study aims to predict air quality in Pekanbaru using the Long Short Term Memory (LSTM) algorithm to develop a time-series air quality prediction model. The algorithm was selected for its ability to achieve high accuracy and low evaluation errors. A total of 2,190 daily data points were used. Testing scenarios included data splitting ratios of 80:20 and 90:10, handling missing values using mean, median, and linear interpolation methods, and varying the number of epochs at 50, 100, and 150. Model evaluation was conducted using the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) metrics. The results indicate that the best scenario was achieved using linear interpolation for handling missing values, a 90:10 data split, and 100 epochs. The resulting prediction model yielded low RMSE and MAPE values, reflecting a satisfactory level of accuracy.

Keywords: Air quality, Prediction, Long Short Term Memory, Air pollution, Pekanbaru.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

KATA PENGANTAR

Assalamu'alaikum wa rohmatullahi wa barokatuh.

Alhamdulillah robbil alamin, tak henti-hentinya kami ucapkan kehadiran Allah *Subhanahu wa ta'ala*, yang dengan rahmat dan hidayah-Nya kami mampu menyelesaikan Tugas Akhir ini dengan baik. Tidak lupa bershalawat kepada Nabi dan Rasul-Nya, Nabi Muhammad *Shallallahu 'alaihi wa salam*, yang telah membimbing kita sebagai umatnya menuju jalan kebaikan.

Tugas Akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar sarjana pada jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Banyak sekali pihak yang telah membantu kami dalam penyusunan laporan ini, baik berupa bantuan materi maupun berupa motivasi dan dukungan kepada kami. Semua itu tentu terlalu banyak bagi kami untuk membalasnya, namun pada kesempatan ini kami hanya dapat mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Hairunas, M. Ag selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M. Pd selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Bapak Iwan Iskandar, M. T selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Bapak Dr. Alwis Nazir, M. Kom selaku Penasehat Akademik sekaligus penguji 1 Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
5. Ibu Dr. Lestari Handayani, S.T., M. Kom selaku Pembimbing I Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
6. Ibu Iis Afrianty, S. T., M. Sc selaku Pembimbing II Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
7. Ibu Fitri Insani, S.T., M. Kom selaku penguji 2 Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
8. Ibu Fadhillah Syafria, S.T., M. Kom selaku Koordinator Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
9. Bapak dan ibu dosen Jurusan Teknik Informatika yang telah memberikan ilmu dan nasehat yang bermanfaat untuk saya.
10. Orang tua yaitu Bapak Sasra dan Ibu Marjuwita, serta adik-adik yaitu Fadil dan Keyla yang selalu memberikan nasehat, dukungan, do'a, finansial dan motivasi saya dalam menyelesaikan tugas akhir ini.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

11. Kepada teman-teman saya kelas Teknik Informatika H yang saling membantu dalam menyelesaikan tugas akhir ini.
12. Dan seluruh pihak terkait yang belum dicantumkan, terima kasih atas dukungannya.

Kami menyadari bahwa laporan ini masih memiliki banyak kekurangan dan kesalahan. Oleh karena itu, kami sangat mengharapkan kritik dan saran yang membangun untuk menyempurnakan laporan ini. Kami juga berharap laporan ini dapat memberikan manfaat bagi setiap pembacanya.

Wassalamu 'alaikum wa rohmatullohi wabarokatuh.

Pekanbaru, 30 Mei 2024

Penulis



UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL.....	iii
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK	vii
ABSTRACT.....	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xiv
DAFTAR RUMUS	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB 2 KAJIAN PUSTAKA.....	4
2.1 Kualitas Udara	4
2.2 Data Mining.....	5
2.3 Prediksi	15

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.4	Penelitian Terkait	15
BAB 3 METODOLOGI PENELITIAN.....		21
3.1	Pengumpulan Data	21
3.1.1	Studi Literatur	21
3.1.2	<i>Dataset</i>	21
3.2	<i>Pre-Processing</i>	22
3.2.1	<i>Cleaning data</i>	22
3.2.2	Penanganan <i>Outlier</i>	22
3.2.3	Normalisasi data.....	23
3.3	Pemodelan <i>Long Short Term Memory (LSTM)</i>	23
3.4	Evaluasi Model.....	23
3.5	Skenario Pengujian.....	24
BAB 4 PEMBAHASAN.....		25
4.1	Pengumpulan Data	25
4.2	<i>Pre-processing</i>	27
4.3	Pemodelan LSTM.....	30
4.4	Hasil Eksperimen	40
4.4.1	Hasil Prediksi	40
4.4.2	Skenario Pengujian.....	42
4.5	Kesimpulan Pengujian.....	45
BAB 5 PENUTUP		47
5.1	KESIMPULAN	47
5.2	SARAN	47
DAFTAR PUSTAKA		48
DAFTAR RIWAYAT HIDUP.....		54

DAFTAR GAMBAR

Gambar 2. 1 Tahapan Data Mining.....	6
Gambar 2. 2 Struktur LSTM	11
Gambar 3. 1 Tahapan Penelitian	21
Gambar 3. 2 Alur Pemrosesan LSTM.....	23
Gambar 4. 1 Data Kualitas Udara Mentah	25
Gambar 4. 2 Grafik Data Kualitas Udara Tahun 2018 Sampai 2023.....	26
Gambar 4. 3 Jumlah <i>Missing Value</i>	27
Gambar 4. 4 Data <i>Outlier</i>	29
Gambar 4. 5 Grafik Data Setelah <i>Preprocessing</i>	29
Gambar 4. 6 Hasil <i>Min-Max Scaling</i> PM ₁₀	31
Gambar 4. 7 Arsitektur LSTM.....	32
Gambar 4. 8 <i>Train Loos</i> dan <i>Validation Loos</i>	34
Gambar 4. 9 Perbandingan data Aktual Dan Prediksi PM ₁₀	37
Gambar 4. 10 Grafik Kualitas Udara 2024	38
Gambar 4. 11 Perbandingan data aktual dan prediksi PM ₁₀ 2024	40
Gambar 4. 12 Perbandingan data aktual dan prediksi SO ₂ 2024.....	40
Gambar 4. 13 Perbandingan data aktual dan prediksi CO 2024	41
Gambar 4. 14 Perbandingan data aktual dan prediksi O ₃ 2024.....	41
Gambar 4. 15 Perbandingan data aktual dan prediksi NO ₂ 2024.....	41
Gambar 4. 16 Perbandingan data aktual dan prediksi PM _{2,5} 2024.....	41

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR TABEL

Tabel 2.1 Tabel Parameter Baku Mutu Udara Ambien Nasional	4
Tabel 2.2 Akurasi Hasil Berdasarkan MAPE	14
Tabel 2.3 Penelitian Terkait	16
Tabel 3.1 Skenario Pengujian	24
Tabel 4.1 Data Kualitas Udara Tahun 2018-2023	26
Tabel 4.2 Dataset Setelah <i>Interpolasi Linear</i>	27
Tabel 4.3 Data PM ₁₀	30
Tabel 4.4 Data validasi Dan Prediksi PM ₁₀	35
Tabel 4.5 Evaluasi Model PM ₁₀	36
Tabel 4.6 Evaluasi Model masing-masing Parameter.....	36
Tabel 4.7 Data Kualitas Udara Tahun 2024.....	37
Tabel 4.8 Prediksi PM ₁₀ 2024	39
Tabel 4.9 Evaluasi Model PM ₁₀ Tahun 2024.....	39
Tabel 4.10 Evaluasi Model masing-masing Parameter Tahun 2024.....	39
Tabel 4.11 Skenario Perbandingan Pada <i>Mean</i>	42
Tabel 4.12 Skenario Perbandingan Pada <i>Median</i>	43
Tabel 4.13 Skenario Perbandingan Pada <i>Interpolasi Linear</i>	44
Tabel 4.14 Perbandingan <i>Mean, Median, dan Interpolasi Linear</i>	45

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR RUMUS

(1) Rumus <i>Mean</i>	7
(2) Rumus <i>Median Ganjil</i>	8
(3) Rumus <i>Median Genap</i>	8
(4) Rumus <i>Interpolasi Linear</i>	8
(5) Rumus <i>Kuartil Bawah (Q1)</i>	9
(6) Rumus <i>Kuartil Atas (Q3)</i>	9
(7) Rumus <i>Interquartile Range (IQR)</i>	9
(8) Rumus <i>Batas Atas</i>	9
(9) Rumus <i>Batas bawah</i>	9
(10) Rumus <i>Min-Max Scaling</i>	10
(11) Rumus <i>Forget Gates LSTM</i>	12
(12) Rumus <i>Input Gates LSTM</i>	12
(13) Rumus menyimpan nilai baru di <i>Cell Memory LSTM</i>	12
(14) Rumus <i>Cell State LSTM</i>	13
(15) Rumus <i>Output Gates LSTM</i>	13
(16) Rumus Nilai prediksi LSTM	13
(17) Rumus Mean Absolute Percentage Errors (MAPE).....	14
(18) Rumus <i>Root Mean Square Error (RMSE)</i>	15

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kualitas udara adalah kondisi atau tingkat kemurnian udara di suatu daerah tertentu, yang menunjukkan seberapa bersih atau tercemar udara tersebut oleh berbagai zat polutan (Santoso et al., 2022). Pencemaran udara adalah pelepasan polutan ke atmosfer yang membahayakan kesehatan manusia dan lingkungan secara keseluruhan (Mustafa et al., 2021). Menurut Peraturan Pemerintah Republik Indonesia nomor 41 tahun 1999, parameter yang mempengaruhi pencemaran udara antara lain Sulfur Dioksida (SO₂), Karbon Monoksida (CO), Nitrogen Dioksida (NO₂), Ozon (O₃), Hidrokarbon (HC), Partikulat Materi (PM₁₀ dan PM_{2,5}), Timah Hitam (Pb), Debu (TSP), dan Dustfall (debu jatuh) (RI, 1999) . Data dari World Health Organization (WHO) menunjukkan bahwa pada tahun 2016 sekitar 6,5 juta orang meninggal akibat pencemaran udara (Handayani et al., 2020).

Pekanbaru yang merupakan salah satu kota besar di Indonesia, menghadapi masalah pencemaran udara. Dinas Lingkungan Hidup kota Pekanbaru melaporkan bahwa kualitas udara Pekanbaru sudah tercemar, terutama akibat kebakaran hutan dan lahan serta aktivitas transportasi. (Insani & Sari, 2020). Pada tahun 2018 hingga 2019 terjadi kebakaran hutan yang mengakibatkan polusi udara di kota Pekanbaru hal ini berdampak bagi aktivitas masyarakat (Limbong et al., 2023). Beberapa masalah yang disebabkan oleh pencemaran udara antara lain penyakit kardiovaskular, kanker paru-paru, infeksi saluran pernapasan akut, asma, dan efek buruk pada hasil kelahiran (Mustafa et al., 2021). Hal ini menjadi perhatian bagi kita semua tentang pentingnya menjaga kualitas udara agar terhindar dari kehidupan yang tidak sehat.

Permasalahan udara yang terjadi di kota Pekanbaru dapat diantisipasi dengan melakukan prediksi kualitas udara. *Algoritma Long Short Term Memory*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

(LSTM) merupakan salah satu algoritma yang dapat digunakan untuk membuat model prediksi data time series kualitas udara, menghasilkan akurasi yang tinggi dan error evaluasi yang rendah (Arkadia et al., 2022). Beberapa penelitian terkait prediksi kualitas udara menggunakan algoritma LSTM, seperti penelitian yang dilakukan oleh (Oktaviani & Hustinawati, 2021) yang membahas tentang Prediksi Rata-Rata Zat Berbahaya di DKI Jakarta Berdasarkan Indeks Standar Pencemar Udara Menggunakan Metode Long Short-Term Memory. Hasil penelitian ini menunjukkan nilai MAPE pada prediksi data training sebesar 12,28%, yang masuk dalam kategori akurat. Penelitian lainnya dilakukan oleh (Ernico et al., 2023) yang meneliti tentang peramalan kualitas udara di Kota Jakarta Pusat dengan metode Long Short-Term Memory dan Support-Vector Regression. Hasil penelitian ini menunjukkan bahwa rata-rata MAPE dari LSTM sebesar 12,15% dan rata-rata RMSE sebesar 0,0941, sedangkan metode SVR memiliki rata-rata MAPE sebesar 16,19% dan rata-rata RMSE sebesar 0,1666. Penelitian lain juga dilakukan oleh (Sundari, 2020) tentang Pemodelan Time Series untuk Peramalan Suhu Udara Menggunakan Metode Long Short Term Memory (LSTM) (Studi Kasus: Stasiun Klimatologi Lampung). Hasil penelitian ini menunjukkan bahwa model LSTM terbaik untuk meramalkan suhu udara minimum terdiri dari lima puluh neuron tersembunyi, empat ukuran batch, dan lima puluh epoch, dengan nilai RMSE sebesar 0,89 dan MAPE sebesar 2,59%.

Dari penelitian tersebut menunjukkan bahwa metode LSTM cocok digunakan untuk memprediksi kualitas udara, dengan akurasi yang tinggi dan tingkat error yang rendah. Oleh karena itu, penulis tertarik menggunakan metode LSTM untuk memprediksi kualitas udara di kota Pekanbaru. Oleh karena itu penulis mengangkat judul "**Prediksi Kualitas Udara Di Kota Pekanbaru Menggunakan Algoritma Long Short Term Memory (LSTM)**". Diharapkan penelitian ini akan menghasilkan prediksi kualitas udara yang akurat dan relevan. Ini juga akan menjadi dasar untuk penelitian serupa di kota-kota lain yang mengalami masalah kualitas udara.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang yang ada, maka didapatkan rumusan masalah dari penelitian ini yakni bagaimana menerapkan algoritma LSTM untuk memprediksi kualitas udara di kota Pekanbaru?

1.3 Batasan Masalah

Penelitian ini membatasi masalahnya dengan membatasi cakupan penelitian ke bidang tertentu yang akan menjadi fokus penelitian ini.

1. Data kualitas udara diperoleh dari Dinas Lingkungan Hidup dan Kebersihan.
2. Penelitian ini dilakukan berdasarkan data pola harian dari kualitas udara kota Pekanbaru dengan parameter PM₁₀, PM_{2.5}, SO₂, CO, O₃, dan NO₂.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menerapkan algoritma *Long Short Term Memory* dalam memprediksi kualitas udara di kota Pekanbaru berdasarkan data yang telah diperoleh.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah memberikan informasi terkait prediksi kualitas udara kepada masyarakat umum dan menambah penelitian tentang kualitas udara.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 2

KAJIAN PUSTAKA

2.1 Kualitas Udara

Kualitas udara adalah kondisi atau tingkat kemurnian udara di suatu daerah atau lokasi tertentu, yang mengacu pada seberapa bersih atau tercemar oleh berbagai zat polutan udara. Kualitas udara menjadi faktor penting yang mempengaruhi kesehatan tenaga kerja (Santoso et al., 2022). Kualitas udara diukur berdasarkan konsentrasi parameter pencemar udara yang melebihi atau berada di bawah nilai baku mutu udara ambien. Baku mutu udara ambien adalah batas atau tingkat maksimal unsur pencemaran udara yang diperbolehkan dalam udara ambien (Subagiyo et al., 2021). Berikut adalah Tabel 2.1 tentang baku mutu ambien nasional menurut Peraturan Pemerintah No. 22 tahun 2021 (RI, 2021):

Tabel 2.1 Tabel Parameter Baku Mutu Udara Ambien Nasional

No	Parameter	Waktu pengukuran	Baku mutu	System pengukuran
1	Sulfur Dioksida (SO ₂)	1 jam	150 µg/m ³	Aktif kontinu
		24 jam	75 µg/m ³	Aktif manual
		1 tahun	45 µg/m ³	Aktif kontinu
2	Karbon Monoksida (CO)	1 jam	10000 µg/m ³	Aktif kontinu
		8 jam	4000 µg/m ³	Aktif kontinu
3	Nitrogen Dioksida (NO ₂)	1 jam	200 µg/m ³	Aktif kontinu
		24 jam	65 µg/m ³	Aktif kontinu
		1 tahun	50 µg/m ³	Aktif kontinu
4	Ozon (O ₃)	1 jam	150 µg/m ³	Aktif kontinu Atif manual
		8 jam	100 µg/m ³	Aktif kontinu
		1 tahun	35 µg/m ³	Aktif kontinu

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5	Hidrokarbon Non 3 jam Metana (NMHC)	3 jam	160 $\mu\text{g}/\text{m}^3$	Aktif kontinu
6	Partikulat Debu 100 Pm (TSP)	24 jam	230 $\mu\text{g}/\text{m}^3$	Aktif manual
		Partikulat Debu < 10 Pm (PM ₁₀)	24 jam	75 $\mu\text{g}/\text{m}^3$
	Partikulat Debu < 2,5 Pm (PM _{2.5})	1 tahun	40 $\mu\text{g}/\text{m}^3$	Aktif manual
		24 jam	55 $\mu\text{g}/\text{m}^3$	Aktif kontinu
1 tahun	15 $\mu\text{g}/\text{m}^3$	Aktif kontinu		
7	Timbal (Pb)	24 jam	2 $\mu\text{g}/\text{m}^3$	Aktif manual

Keterangan $\mu\text{g}/\text{m}^3$ = konsentrasi dalam mikrogram per meter kubik.

Pada tabel diatas menunjukkan pengukuran dari masing-masing parameter yang mempengaruhi kualitas udara, dengan demikian jika nilai dari parameter tersebut melebihi standar dari baku mutu ambien maka itu bisa menunjukkan adanya pencemaran udara. Pencemaran udara merujuk pada pelepasan polutan ke atmosfer yang berdampak buruk pada kesehatan manusia dan planet secara keseluruhan (Mustafa et al., 2021). Tiga kategori utama sumber pencemaran udara adalah emisi alami, sumber perkotaan dan industri, sumber pedesaan dan pertanian, dan lainnya. Sumber industri utama biasanya berasal dari titik, tetapi juga ada sumber garis, seperti jalan penghubung di daerah tersebut (Novelan, 2020). Pencemar udara menyebabkan banyak munculnya penyakit seperti asma, bronkitis sehingga menimbulkan lingkungan yang tidak sehat.

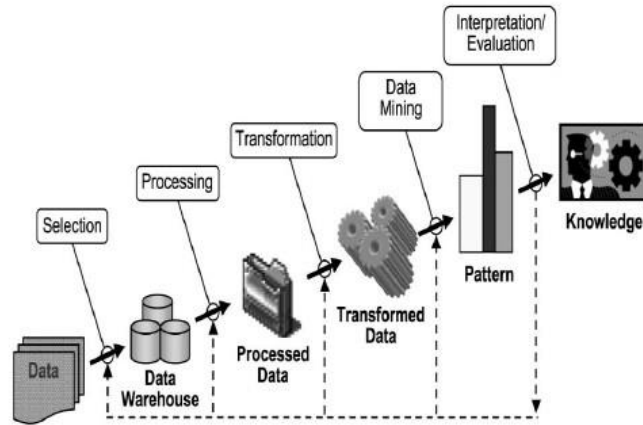
2.2 Data Mining

Data mining adalah suatu proses penggalian data besar yang belum diketahui sebelumnya. Berdasarkan tugasnya, data mining dibagi menjadi asosiasi, pengklasteran (clustering), prediksi, estimasi, dan klasifikasi (Akram et al., 2021). Selain itu data mining juga dijelaskan sebagai proses teknik matematika, statistik, dan *machine learning* yang digunakan untuk mengidentifikasi informasi

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

yang berguna (Rofiq et al., 2020). Berikut adalah 5 tahapan data mining yang biasa disebut dengan KDD (*Knowledge Discovery Database*) dapat dilihat pada gambar di bawah ini:



Gambar 2. 1 Tahapan Data Mining

2.2.1.1 Selection

Sebelum memulai penggalian informasi dalam proses KDD, dilakukan pemilihan atau seleksi data dari kumpulan data operasional. Data yang dipilih untuk proses data mining disimpan dalam berkas terpisah dari basis data operasional. Sebelum proses data mining dimulai, dilakukan tahap pengembangan dengan menggabungkan data yang ada dengan data atau informasi relevan lainnya yang diperlukan untuk KDD, termasuk data atau informasi dari sumber eksternal (Bahtiar, 2023).

2.2.1.2 Preprocessing

Pada *preprocessing* data dilakukan dua tahap *preprocessing* yaitu dengan melakukan *cleaning* data untuk menghilangkan *missing value* dan penanganan outlier.

A. Cleaning data

Setelah proses pengumpulan data dilakukan maka selanjutnya melakukan proses *cleaning*. Proses *cleaning* ini mencakup penghapusan duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data. Pada proses ini juga dilakukan proses *enrichment*, yakni proses “memperkaya” data yang

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

sudah ada dengan informasi lain yang relevan dan diperlukan (Yuli Mardi, 2019). Proses *cleaning* ini juga menangani kesalahan pada hilangnya data atau yang dikenal dengan *missing value* yang akan menyebabkan data menjadi tidak lengkap.

Missing value menyebabkan beberapa masalah yang dapat menyebabkan hasil dari penelitian menjadi kurang tepat. Ada empat masalah yang akan disebabkan oleh *missing value* yaitu, hilangnya data mengurangi kekuatan statistik, yang berarti bahwa pengujian akan menolak hipotesis nol ketika hipotesis itu salah. Kedua, dapat mengakibatkan pendugaan parameter menjadi tidak akurat. Ketiga, dapat menurunkan representasi sampel. Keempat, analisis penelitian mungkin menjadi lebih sulit (Fadila & Muchlisoh, 2022). Oleh karena itu, penanganan *missing value* diperlukan untuk dilakukan.

Penanganan *missing value* dapat dilakukan dengan dua cara yakni menghapus seluruh baris pada data yang hilang dan menghitung nilai pengganti pada nilai yang kosong dengan nilai konstanta, nilai rata-rata atau *mean*, *median*, atau metode imputasi lainnya (Sundari, 2020). Adapun persamaan dari nilai *mean*, *median*, dan *interpolasi linear* dari suatu data dapat dirumuskan persamaan berikut:

1) *Mean*

Berikut ini adalah rumusnya:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} \text{ atau } \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Keterangan:

\bar{X} = Rata-rata

$\sum_{i=1}^n x_i$ = Jumlah seluruh nilai data

n = Jumlah seluruh frekuensi

2) *Median*

Rumus *Median* terbagi menjadi dua yaitu *Median* untuk nilai ganjil dan *Median* untuk nilai genap. Berikut ini adalah rumusnya:

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Jika n ganjil:

$$\tilde{x} = x \frac{n+1}{2} \tag{2}$$

- Jika n genap:

$$\tilde{x} = \frac{x \frac{n}{2} + x \frac{n}{2} + 1}{2} \tag{3}$$

Keterangan:

- x = Kumpulan data yang sedang dihitung *Median*
- \tilde{x} = Nilai *Median* dari Kumpulan data
- n = Jumlah total data dalam kumpulan data
- x_i = Nilai ke-I dalam Kumpulan data

3) *Interpolasi Linear*

Berikut ini adalah rumusnya:

$$f_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{(x_1 - x_0)}(x - x_0) \tag{4}$$

Keterangan:

- $f_1(x)$ = Nilai fungsi yang diinterpolasi pada titik x
- $f(x_0)$ = Nilai fungsi pada titik data pertama
- $f(x_1)$ = Nilai fungsi pada titik data kedua
- x_0 dan x_1 = Dua titik data yang diketahui, dengan $x_0 < x_1$
- x = Titik yang ingin diperkirakan nilainya, dengan $x_0 \leq x \leq x_1$

B. Penanganan Outlier

Kondisi outlier terjadi ketika nilai-nilai parameter yang sangat ekstrim muncul dalam suatu dataset. Keberadaan outlier dapat menyebabkan kesalahan dalam klasifikasi model, memberikan bias dalam estimasi parameter, menghasilkan hasil yang tidak akurat, serta menurunkan kualitas prediksi (Siringoringo et al., 2022).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Ada beberapa faktor yang menyebabkan munculnya data outlier, seperti kesalahan dalam proses entri data, gagal dalam mengidentifikasi *missing value* pada program komputer, outlier berasal dari data yang tidak termasuk dalam populasi sampel, outlier berasal dari populasi sampel yang diambil, namun distribusi variabel dalam populasi tersebut memiliki nilai ekstrim dan tidak mengikuti distribusi normal (Pradana & Astika, 2019).

Adapun cara mengatasi *outlier* dapat dilakukan dengan menghapusnya, namun pada umumnya outliers dapat diatasi menggunakan *metric* IQR (*Interquartile Range*). Berikut langkah-langkah tahap *outliers removal*:

1. Menghitung nilai Q3 (Kuartil atas) dan nilai Q1 (Kuartil bawah) pada data. Berikut:

$$Q1 = \frac{25}{100}(n) \tag{5}$$

$$Q3 = \frac{75}{100}(n) \tag{6}$$

2. Menghitung nilai IQR dengan cara selisih antara Q3 dengan Q1. Berikut persamaan dari IQR:

$$IQR = Q3 - Q1 \tag{7}$$

3. Menentukan nilai batas atas dan batas bawah

$$Batas\ atas = Q3 + 1.5 * IQR \tag{8}$$

$$Batas\ bawah = Q1 - 1.5 * IQR \tag{9}$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. Membandingkan nilai data dengan batas atas dan batas bawah: Apabila suatu nilai data terletak di luar batas tersebut, maka nilai tersebut dapat dianggap sebagai *outlier*.

Keterangan:

IQR = Nilai *interquartile range*

Q3 = Nilai Kuartil 3

Q1 = Nilai *Kuartil* 1 (Irianto et al., 2022).

2.2.1.3 Transformasi

Sebelum dapat digunakan, metode data mining biasanya membutuhkan format atau struktur data tertentu. Mengubah data yang ada dari satu format atau struktur ke format atau struktur lain yang siap diolah adalah proses yang dikenal sebagai transformasi. Proses ini memungkinkan pengolahan data yang lebih efisien. Pola yang ditemukan juga lebih mudah dipahami (Loelianto et al., 2020). Proses tranformasi ini bisa mencakup normalisasi data, konversi format, atau pembentukan fitur baru yang lebih relevan untuk analisis (Rahayu et al., 2018).

Normalisasi bertujuan mengubah nilai variabel agar memiliki skala yang sama, sehingga variabel dengan rentang besar tidak mendominasi perhitungan. Dengan skala seragam, semua variabel berkontribusi secara adil dalam analisis atau model (Rahayu et al., 2018). Terdapat banyak teknik normalisasi seperti *Min-Max*, *Z score* dan *Decimal Scaling* (Riska Chairunisa et al., 2020).

Min-max normalization adalah metode transformasi data linear yang menyesuaikan nilai ke dalam rentang yang sama, biasanya [0, 1]. Metode ini mempercepat konvergensi dibandingkan metode lainnya (Gde Agung Brahmama Suryanegara et al., 2021). Teknik *Min-Max Scaling* dapat dirumuskan ke dalam persamaan berikut:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{10}$$

x' = Data kualitas udara yang telah dinormalisasikan.

Hak Cipta Dilindungi Undang-Undang

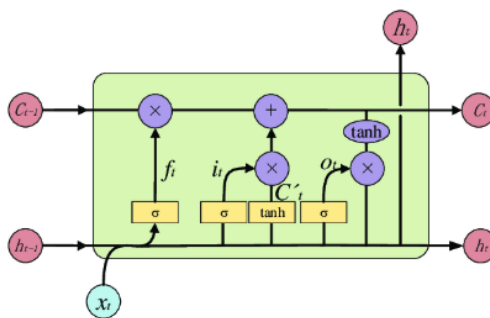
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- x = Data aktual Kualitas udara.
- x_{min} = Nilai minimum data kualitas udara.
- x_{max} = Nilai maximum data kualitas udara (Sundari, 2020).

2.2.1.4 Data Mining (Long Short Term Memory)

Long Short Term Memory (LSTM) adalah jaringan yang dirancang untuk menangani masalah ketergantungan jangka panjang dalam data berurutan. LSTM banyak digunakan dalam analisis pola seperti pemodelan bahasa dan akustik skala besar. Keunggulan utama LSTM adalah kemampuannya untuk mempertahankan informasi jangka panjang melalui struktur gate yang dapat menambah atau menghapus informasi dalam status sel, berbeda dengan RNN yang hanya memiliki modul berulang sederhana. (Arkadia et al., 2022).

Dalam arsitektur LSTM, sel memori memiliki tiga gate: *input*, *output*, dan *forget*. Gate input mengatur seberapa banyak nilai baru masuk ke sel, gate *forget* mengatur seberapa banyak nilai yang disimpan, dan gate output mengatur seberapa banyak nilai yang digunakan untuk keluaran LSTM. Sel LSTM menyimpan masukan untuk waktu yang lama, mengelola informasi melalui ketiga gate ini. (Khumaidi et al., 2020). Gambar 2.2 menunjukkan struktur dari algoritma LSTM, berikut gambarnya:



Sumber: (Hidayatullah et al., 2022)

Gambar 2. 2 Struktur LSTM

Data yang masuk ke *forget gates* akan diproses sesuai dengan informasinya, dan sejumlah data akan dipilih untuk disimpan di *cell memory*. Prinsip kerja

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

sigmoid untuk fungsi aktivasinya dijelaskan dalam persamaan berikut (Hidayatullah et al., 2022):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (11)$$

Keterangan:

- f_t = Forget gate
- σ = Fungsi sigmoid
- W_f = Nilai Weight untuk forget gate
- h_{t-1} = Nilai *output* sebelum orde ke-t
- x_t = Nilai *input* sebelum orde ke-t
- b_f = Nilai bias pada *forget gate*

Sedangkan 2 gate yang terhubung ke *input* memiliki fungsi aktivasi *sigmoid*, yang memperbarui informasi, dan fungsi aktivasi *tanh*, yang menyimpan nilai baru di *cell memory*. Ini dapat dilihat dari persamaan berikut:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (12)$$

Keterangan:

- i_t = Input gate
- W_i = Nilai Weight untuk forget gate
- b_i = Nilai bias pada *input gate*

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (13)$$

Keterangan:

- \tilde{C}_t = Nilai baru yang ditambahkan ke *cell state*

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$Tanh$ = fungsi tanh

W_c = Nilai Weight untuk *cell state*

b_c = Nilai bias pada *cell state*

Nilai *input gate* akan digabungkan dengan nilai *cell memory*, seperti yang ditunjukkan dalam persamaan berikut:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (14)$$

Keterangan:

C_{t-1} = *Cell state* sebelum orde ke-t

Ada 2 gate pada *output gates*. Fungsi aktivasi *sigmoid* dapat memilih nilai yang akan dikeluarkan dan fungsi aktivasi *tanh* dapat menyimpan nilai. Ini dapat dilihat dari persamaan berikut:

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (15)$$

Keterangan:

O_t = *Output gate*

W_o = Nilai Weight untuk *output gate*

b_o = Nilai bias pada *output gate*

$$h_t = O_t \tanh(C_t) \quad (16)$$

Keterangan:

h_t = Nilai *output* orde ke-t

C_t = *Cell state*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.2.1.5 Interpretation/ Evaluation

Pola informasi yang dihasilkan dari proses data mining harus disajikan dalam format yang mudah dipahami oleh semua pihak yang terlibat. Tahap ini merupakan bagian dari proses Interpretasi (KDD). Pada tahap ini, dilakukan pemeriksaan untuk memastikan bahwa pola atau informasi yang ditemukan tidak bertentangan dengan fakta atau hipotesis sebelumnya (Yuli Mardi, 2019). Beberapa matriks yang digunakan untuk mengevaluasi kinerja model prediksi sebagai berikut:

1. *Mean Absolute Percentage Errors* (MAPE)

Pengukuran kesalahan relatif dikenal sebagai *Mean Absolute Percentage Errors* (MAPE). MAPE memberikan informasi persentase kesalahan terlalu tinggi atau terlalu rendah dan menunjukkan persentase kesalahan hasil peramalan terhadap permintaan aktual selama periode tertentu (Ernico et al., 2023). Persamaan MAPE adalah sebagai berikut:

$$MAPE = \frac{\sum \frac{|X_i - F_i|}{X_i} 100\%}{N} \tag{17}$$

Keterangan:

- X_i = Nilai aktual
- F_i = Nilai peramalan
- N = Jumlah data

Tabel 2.2 Kategori Range Nilai MAPE

MAPE	Kategori
<10%	Sangat Baik
10-20%	Baik
20-50%	Layak/Cukup
>50%	Buruk

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.2 menunjukkan rentang akurasi dari evaluasi model menggunakan MAPE (Sundari, 2020).

2. *Root Square Error* (RMSE)

Root Square Error (RMSE) adalah akar kuadrat estimasi perbedaan nilai peramalan dengan nilai aktual. Nilai RMSE yang lebih tinggi menunjukkan bahwa model peramalan yang dihasilkan lebih akurat (Ernico et al., 2023). Persamaan RMSE adalah sebagai berikut:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(P_m - P_r)^2}{N}} \tag{18}$$

Keterangan:

- P_m = Nilai aktual
- P_r = Nilai peramalan
- N = Jumlah data

2.3 Prediksi

Prediksi adalah proses perhitungan sistematis yang digunakan untuk memperkirakan kemungkinan suatu objek di masa depan berdasarkan data dari masa lalu dan masa kini (Oktaviani & Hustinawati, 2021). Prediksi juga merupakan hasil dari memproyeksikan nilai ke masa depan berdasarkan informasi dari masa lalu. Prediksi ini membantu dalam perencanaan dan pengambilan keputusan dengan memprediksi apa yang akan terjadi dalam situasi tertentu (Farikhul Firdaus & Papatungan, 2022).

2.4 Penelitian Terkait

Penelitian penting untuk memanfaatkan hasil penelitian sebelumnya yang relevan, yang merujuk pada temuan penelitian sebelumnya yang berkaitan dengan topik penelitian yang sedang dilakukan. Hasil dan diskusi dari penelitian

sebelumnya yang berkaitan dengan topik penelitian yang sedang dilakukan disajikan sebagai berikut:

Tabel 2.3 Penelitian Terkait

No.	Nama Peneliti	Tahun	Judul Penelitian	Hasil Penelitian
1	Mohammad Ernico Suryo Wicaksono, Gusti Made Arya Sasmitaa, I Putu Agus Eka Pratama (Ernico et al., 2023)	2023	Peramalan Kualitas Udara Di Kota Jakarta Pusat Dengan Metode <i>Long Short-Term Memory</i> Dan <i>Support-Vector Regression</i>	Rata-rata MAPE dari LSTM sebesar 12,15% dan rata-rata RMSE sebesar 0,0941, sedangkan Metode SVR memiliki rata-rata MAPE sebesar 16,19% dan rata-rata RMSE sebesar 0,1666.
2	Linda Sundari (Sundari, 2020)	2023	Pemodelan Time Series Untuk Peramalan Suhu Udara Menggunakan Metode <i>Long Short Term Memory</i> (LSTM) (Studi Kasus: Stasiun Klimatologi Lampung)	Model LSTM terbaik untuk meramalkan suhu udara minimum terdiri dari lima puluh neuron tersembunyi, empat ukuran batch, dan lima puluh epoch, dan memiliki nilai RMSE 0,89 dan MAPE 2,59%.
3	Adli A Nababan, Miftahul Jannah, Mia	2023	Prediksi Kualitas Udara Menggunakan Xgboost Dengan	Nilai rata-rata akurasi sebesar 98%, nilai precision sebesar 79%, nilai recall sebesar 79%,

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	Aulina, Dwiki Andrian (Nababan et al., 2023)		<i>Synthetic Minority Oversampling Technique (SMOTE)</i> Berdasarkan Indeks Standar Pencemaran Udara (ISPU)	nilai F1-score sebesar 98%, dan nilai AUC-ROC sebesar 99%
4	Arvi Arkadia, Bayu Hananto, Desta Sandya Prasvita (Arkadia et al., 2022)	2022	Optimasi <i>Long Short Term Memory</i> Dengan Adam Menggunakan Data Udara Kota DKI Jakarta	Nilai R2 untuk pelatihan dan pengujian masing- masing sebesar 0,632 dan 0,570 menunjukkan bahwa model LSTM cocok digunakan dalam prediksi PM _{2,5} .
5	Yadi Karyadi, Handri Santoso Pradita (Karyadi, 2022)	2022	Prediksi Kualitas Udara Dengan Metoda LSTM, <i>Bidirectional</i> LSTM, dan GRU	Model LSTM dan LSTM <i>Bidirectional</i> lebih baik daripada model GRU dalam memprediksi kualitas udara berdasarkan parameter yang ada.
6	Riza Farikhul Firdaus (Farikhul Firdaus & Paputungan, 2022)	2022	Prediksi Curah Hujan Menggunakan Metode <i>Long Short Term Memory</i> (Studi	Nilai Train Score RMSE sebesar 12.21 dan Test Score RMSE sebesar 8.78.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

			Kasus: Kota Bandung)	
7	Benny Khalid Hidayatullah, Meta Kallista, Casi Setianingsih (Hidayatullah et al., 2022)	2022	Prediksi Indeks Standar Pencemar Udara Menggunakan Metode <i>Long Short-Term Memory</i> Berbasis Web (Studi Kasus Pada Kota Jakarta)	Nilai RMSE dari 0.0431 sampai 0.07237 dan skor R ² dari rentang nilai 0.58157 hingga 0.88621.
8	Ahmad Yusuf, Kusriani, Alva Hendi Muhammad (Yusuf et al., 2022)	2022	Perbandingan <i>Additive</i> Dan <i>Multiplicative Exponential Smoothing</i> Terhadap Prakiraan Kualitas Udara Di Banjarmasin	Model TES <i>Additive</i> dan <i>Multiplicative</i> memiliki hasil prakiraan yang cukup baik dengan nilai error yang rendah dan akurasi yang baik. $\alpha = 0.6$, $\alpha = 0.02$, dan $\alpha = 0.02$ untuk metode TES <i>Additive</i> , sedangkan $\alpha = 0.445$, $\alpha = 0.001$, dan $\alpha = 0.117$.
9	Manda Nurrohman Akuba, Samuel Fernandez Pardede, Adrian Fransisco Maskim,	2022	Perbandingan Model Dalam Memprediksi Karbon Monoksida di Kota Medan	Ditunjukkan bahwa model <i>hybrid</i> yang menggunakan metode <i>Machine Learning Neural Prophet</i> , <i>Auto-Regressive</i> , dan <i>Hybrid</i> memiliki tingkat akurasi yang tinggi dalam

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak cipta milik UIN Suska Riau	Dosso Ganimel Aditya Simanullang (Akuba et al., 2022)		Bulan Januari 2023	memprediksi jumlah CO ₂ dengan nilai RMSE sebesar 3,52 dan nilai MAE sebesar 2,28.
10	Adinda Amalia, Ati Zaidiah, Ika Nurlaili Isnainiyah (Amalia et al., 2022)	2022	Prediksi Kualitas Udara Menggunakan Algoritma <i>K-Nearest Neighbor</i>	Nilai K yang ideal adalah K = 7, dengan akurasi 96%, presisi 92%, recall 95%, dan pengukuran f 93%. Sistem prediksi kualitas udara yang dibangun menggunakan <i>framework Flask</i> dapat menampilkan hasil prediksi secara menyeluruh.
11	Anisa Oktaviani, Hustinawati (Oktaviani & Hustinawati, 2021)	2021	Prediksi Rata-Rata Zat Berbahaya di Dki Jakarta Berdasarkan Indeks Standar Pencemar Udara Menggunakan Metode <i>Long Short-Term Memory</i>	Nilai MAPE pada prediksi data training adalah 12.28%, masuk dalam kategori akurat.
12	Ali Khumaidi, Ridwan	2020	Pengujian Algoritma <i>Long Short Term</i>	Hasil prediksi suhu menunjukkan keakuratan yang cukup baik, hasilnya

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

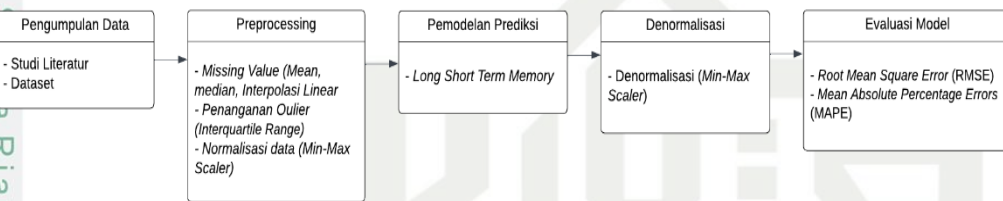
	Raafi'udin, Indra Permana Solihin (Khumaidi et al., 2020)		Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung	menunjukkan RMSE sebesar 1,76 dan standar deviasi uji sebesar 1,75 untuk prediksi PM ₁₀ .
13	Muh. Anas Faisho, Endroyono, dan Astria Nur Irfansyah (Faishol et al., 2020)	2020	Prediksi Polusi Udara Perkotaan di Surabaya Menggunakan <i>Recurrent Neural Network – Long Short Term Memory</i>	Komposisi data latih dan data uji adalah 95%:5% dari beberapa skenario pembagian data latih dan data uji. Nilai sebesar 1,880 diperoleh dengan menggunakan perhitungan error RMSE.
14	Muhammad Idham Habibie (Habibie, 2020)	2020	Analisis Prediksi pada variabel Temperatur dan Kelembaban di sensor IoT menggunakan metode LSTM	Hasil penelitian ini menunjukkan bahwa pemodelan yang digunakan berjalan dengan baik dengan hasil prediksi RMS = 0.032 Loss = 0.001 MAE = 0.007 R2 Square = 0.68
15	Marie Luthfi Ashari, Mujiono Sadiki (Ashari & Sadikin, 2020)	2020	Prediksi Data Transaksi Penjualan Time Series Menggunakan Regresi LSTM	Nilai RMSE dengan menggunakan data 90% dan 10%, interval rentang, metode tiga jendela, dan 100 epoch. Nilai RMSE untuk pelatihan adalah 286.465.424 dan untuk pengujian adalah 187.013.430.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 3

METODOLOGI PENELITIAN

Penelitian ini dilakukan dengan metode terstruktur yang membantu dalam perencanaan, pelaksanaan, dan evaluasi proyek prediksi. Berikut adalah tahapannya:



Gambar 3. 1 Tahapan Penelitian

3.1 Pengumpulan Data

Tahap pengumpulan data bertujuan untuk mengumpulkan informasi yang akan digunakan dalam penelitian. Proses ini dilakukan dengan mengumpulkan teori-teori yang relevan dan berkaitan dengan penelitian ini. Teori-teori yang digunakan dalam penelitian ini berupa teori tentang kualitas udara, data mining, prediksi, algoritma LSTM, evaluasi model, serta penelitian-penelitian yang berkaitan dengan penelitian ini. Pengumpulan data dilakukan dalam dua tahapan yakni:

3.1.1 Studi Literatur

Informasi yang digunakan dalam penelitian ini berasal dari studi literatur yang relevan dengan topik penelitian. Pada tahap ini, sumber tertulis seperti artikel jurnal ilmiah, buku, tesis, laporan penelitian yang berkaitan dengan prediksi kualitas udara dan algoritma LSTM. Dasar pengetahuan yang akan dibangun dalam penelitian ini akan berasal dari literatur ini.

3.1.2 Dataset

Dataset yang akan digunakan dalam penelitian ini merupakan data sekunder. Data sekunder adalah data yang merujuk pada sumber yang telah tersedia

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengummumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

untuk digunakan dalam penelitian. *Dataset* yang digunakan dalam penelitian ini adalah data kualitas udara kota Pekanbaru dari tanggal 1 Januari 2018 sampai 18 Oktober 2024. Data kualitas udara ini diperoleh dari Dinas Lingkungan Hidup dan Kebersihan Kota Pekanbaru yang berlokasi di Jl. Datuk Setia Maharja No. 4, Kelurahan Simpang Tiga, Kecamatan Bukit Raya.

3.2 Pre-Processing

Pada proses *preprocessing* data dilakukan dua tahap *preprocessing* yaitu dengan melakukan *cleaning* data untuk menghilangkan *missing value*, penanganan *outlier*, dan normalisasi data berikut ini adalah tahapnya:

3.2.1 Cleaning data

Pada *Cleaning* data dilakukan proses *selection* data yang bertujuan memilih data dari data kualitas untuk dilakukan analisis lebih lanjut. *Selection* data dilakukan secara manual, adapun data yang di *selection* yakni tanggal/date, serta data dari parameter yang ada yakni PM_{10} , $PM_{2,5}$, SO_2 , CO , O_3 , dan NO_2 .

Penanganan *missing value* juga dilakukan pada proses ini, dengan tujuan menghilangkan nilai yang kosong pada tiap parameter kualitas udara. Penanganan *Missing value* dilakukan untuk memastikan kualitas dan *integritas* data sehingga analisis atau model dapat memberikan hasil yang akurat. Penanganan *missing value* pada penelitian ini dilakukan dengan menggunakan teknik *imputasi mean*, *median*, dan *interpolasi linear* yang didapatkan dari Persamaan (1) sampai Persamaan (4).

3.2.2 Penanganan Outlier

Melakukan penanganan *outlier* yang bertujuan untuk mengatasi nilai-nilai ekstrim yang ada dalam data. Penanganan ini dilakukan agar data menjadi lebih bersih dan *representatif*, sehingga dapat meningkatkan akurasi dalam prediksi. Menghilangkan *outlier* juga membantu mencegah *distorsi* hasil, mengurangi bias, dan memastikan model bekerja lebih optimal dengan data yang lebih konsisten. Penanganan outlier pada penelitian ini dilakukan dengan *metric* IQR dengan Persamaan (5) sampai Persaman (9).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

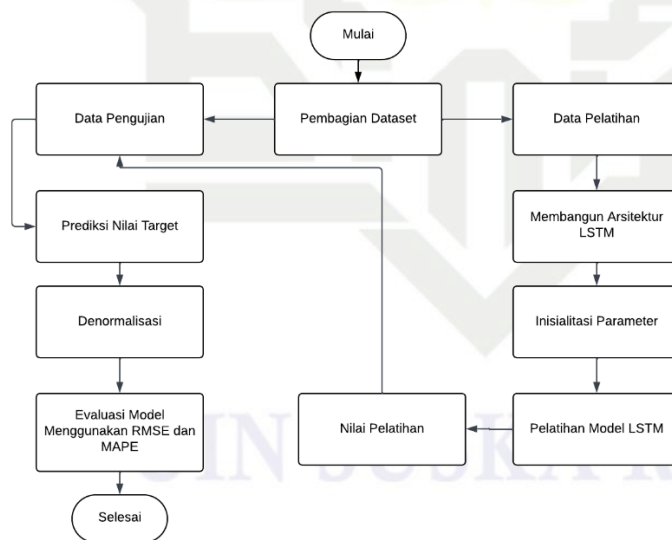
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.2.3 Normalisasi data

Normalisasi data dilakukan untuk menyesuaikan nilai-nilai dalam dataset sehingga skala atau rentang menjadi seragam. Penelitian ini menggunakan metode normalisasi data *min-max scaler* yang biasanya dengan rentang angka 0 dan 1 dengan Persamaan (10).

3.3 Pemodelan *Long Short Term Memory* (LSTM)

Sebelum memulai pelatihan model, dataset dipisah menjadi dua bagian utama yakni data pelatihan dan data uji. Ada dua perbandingan pembagian data pelatihan dan pengujian yang akan digunakan dalam penelitian ini yakni 80:20 dan 90:10 persen. Tujuan dari pemisahan data ini adalah untuk menentukan seberapa akurat model dapat menggeneralisasi dan membuat prediksi dari data yang belum pernah dilihat sebelumnya dan menganalisis pembagian data yang lebih akurat dari ketiga pembagian tersebut. Metode LSTM digunakan dalam mode order, memisahkan data berdasarkan urutan bulan, dan kemudian model LSTM dilatih menggunakan data pelatihan. Berikut alur pelatihan model LSTM:



Gambar 3. 2 Alur Pemrosesan LSTM

3.4 Evaluasi Model

Pada tahapan ini, evaluasi akan dilakukan menggunakan metrik evaluasi RMSE dengan Persamaan (18) dan MAPE dengan Persamaan (17). Hasil dari

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

metrik ini akan memberikan evaluasi serta pemahaman tentang bagaimana model LSTM berfungsi pada dataset yang diuji. Nilai RMSE yang lebih kecil menunjukkan kualitas peramalan yang lebih baik, sebagaimana dijelaskan dalam penelitian Rofiq et al. (2020). Demikian pula, nilai MAPE yang semakin rendah menunjukkan bahwa hasil prediksi semakin mendekati nilai sebenarnya, yang juga menjadi indikator keakuratan model berdasarkan penelitian Sundari (2020). Ini akan menjadi panduan utama dalam mengevaluasi keakuratan model LSTM dalam memprediksi kualitas udara.

3.5 Skenario Pengujian

Pengujian dilakukan dalam beberapa skenario dengan tujuan untuk menemukan model terbaik dalam memprediksi kualitas udara. Skenario pengujian melibatkan pembagian data pelatihan dan pengujian, serta berbagai metode penanganan *missing value* dan jumlah *epoch* yang berbeda. Berikut ini adalah tabel skenarionya:

Tabel 3.1 Skenario Pengujian

Pembagian data	Metode <i>missing value</i>	<i>Epoch</i>
80:20	<i>Mean</i>	50
90:10	<i>Median</i>	100
	<i>Interpolasi Linear</i>	150

Tabel 3.1 menggambarkan skenario pengujian yang dilakukan untuk menguji berbagai kombinasi pembagian data, metode penanganan *missing value*, dan jumlah *epoch* yang berbeda dalam proses pengujian. Pengujian ini bertujuan untuk menentukan pendekatan terbaik dari berbagai skenario yang dilakukan. Ada beberapa skenario pengujian menggunakan pendekatan yang berbeda namun menghasilkan hasil yang bagus. Sebagai contoh, penelitian Ashari & Sadikin (2020), yang menemukan bahwa skenario terbaik menggunakan pembagian data 90:10 dengan epoch 100. Sedangkan penelitian lainnya, seperti yang dilakukan oleh Sundari (2020), yang menemukan skenario terbaiknya dengan pembagian data 80:20 dan epoch 50.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 5

PENUTUP

5.1 KESIMPULAN

Kesimpulan dari penelitian prediksi kualitas udara di kota Pekanbaru menggunakan algoritma LSTM adalah:

1. Algoritma LSTM telah diterapkan dalam memprediksi kualitas udara di kota Pekanbaru.
2. Setelah beberapa percobaan, hasil terbaik diperoleh dengan pembagian data 90-10, 100 *epoch*, dan penanganan *missing value* menggunakan *interpolasi linear*, menghasilkan nilai RMSE dan MAPE yang rendah secara keseluruhan.
3. Faktor cuaca juga mempengaruhi kualitas udara, menyebabkan beberapa parameter tidak seimbang dan menghasilkan perbedaan signifikan antara nilai aktual dan prediksi.

5.2 SARAN

Untuk meningkatkan akurasi dari prediksi pada penelitian selanjutnya disarankan melakukan:

1. Melakukan studi lebih lanjut untuk mengatasi keterbatasan data untuk meningkatkan akurasi prediksi kualitas udara, hal ini merujuk pada penelitian sebelumnya dari Prabowo (2019).

DAFTAR PUSTAKA

- Akram, A., Risal, N., Yusuf, N. I., Kaswar, A. B., & Adiba, F. (2021). Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma Naive Bayes. *Indonesian Journal of Fundamental Sciences*, 7(1), 18–28. <https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/covidwho-1646370>
- Akuba, M. N., Pardede, S. F., Maskim, A. F., & Simanulang, D. G. A. (2022). Perbandingan Model Dalam Memprediksi karbon Monoksida Di Kota Medan Bulan Januari 2023. *DiJITAC*, 2(2), 47–55.
- Amalia, A., Zaidiah, A., & Isnainiyah, I. N. (2022). Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 7(2), 496–507. <https://doi.org/10.29100/jipi.v7i2.2843>
- Arkadia, A., Hananto, B., & Prasvita, D. S. (2022). Optimasi Long Short Term Memory Dengan Adam Menggunakan Data Udara Kota DKI Jakarta. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 92–101.
- Ashari, M. L., & Sadikin, M. (2020). Prediksi Data Transaksi Penjualan Time Series Menggunakan Regresi Lstm. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 9(1), 1. <https://doi.org/10.23887/janapati.v9i1.19140>
- Bahtiar, R. (2023). Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor. *Jurnal Informatika MULTI*, 1(3), 203–214. <https://jurnal.publikasitecno.id/index.php/jim203>
- Ernico, M., Wicaksono, S., Made, G., Sasmita, A., & Eka, I. P. A. (2023). Pusat Dengan Metode Long Short - Term Memory Dan Support - Vector Regression. *Jitter*, 4(1).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Fadila, L. M. A., & Muchlisoh, S. (2022). Perbandingan Kinerja Metode Hybrid KNNI-GA dan MissForest Dalam Menangani Missing Values. *Seminar Nasional Official Statistics*, 2022(1), 553–562. <https://doi.org/10.34123/semnasoffstat.v2022i1.1315>
- Faihsol, M. A., Endroyono, E., & Irfansyah, A. N. (2020). Predict Urban Air Pollution in Surabaya Using Recurrent Neural Network – Long Short Term Memory. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 18(2), 102. <https://doi.org/10.12962/j24068535.v18i2.a988>
- Fariikhul Firdaus, R., & Paputungan, I. V. (2022). Prediksi Curah Hujan di Kota Bandung Menggunakan Metode Long Short Term Memory. In *Jurnal Penelitian Inovatif* (Vol. 2, Issue 3, pp. 453–460). <https://doi.org/10.54082/jupin.99>
- Gde Agung Brahma Suryanegara, Adiwijaya, & Mahendra Dwifebri Purbolaksono. (2021). Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 114–122. <https://doi.org/10.29207/resti.v5i1.2880>
- Habibie, M. I. (2020). Analisis Prediksi pada variabel Temperatur dan Kelembapan di sensor IoT menggunakan metode LSTM. *Seri Prosiding Seminar Nasional Dinamika Informatika* 2020, 51–54. <http://prosiding.senadi.upy.ac.id/index.php/senadi/article/view/125%0Ahttp://prosiding.senadi.upy.ac.id/index.php/senadi/article/viewFile/125/117>
- Handayani, A. S., Soim, S., Agusdi, T. E., Rumiasih, & Nurdin, A. (2020). Klasifikasi Kualitas Udara Dengan Metode Support Vector Machine. *JIRE (Jurnal Informatika & Rekayasa Elektronika)*, 3(2), 187–199.
- Hidayatullah, B. K., Kallista, M., Setianingsih, C., S1, P., & Komputer, T. (2022). Prediksi Indeks Standar Pencemar Udara Menggunakan Metode Long Short-

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Term Memory Berbasis Web (Studi Kasus Pada Kota Jakarta). *E-Proceeding of Engineering*, 9(3), 1247–1255. <https://data.jakarta.go.id/>

Insani, F., & Sari, A. P. (2020). Optimzation of Interval Fuzzy Time Series With Particle Swarm Optimization for Prediction Air Quality on Pekanbaru. *Indonesian Journal of Artificial Intelligence and Data Mining*, 3(1), 36. <https://doi.org/10.24014/ijaidm.v3i1.9298>

Irianto, M. R., Maududie, A., & Arifin, F. N. (2022). Implementation of K-Means Clustering Method for Trend Analysis of Thesis Topics (Case Study: Faculty of Computer Science, University of Jember). *Berkala Sainstek*, 10(4), 210. <https://doi.org/10.19184/bst.v10i4.29524>

Karyadi, Y. (2022). Prediksi Kualitas Udara Dengan Metoda LSTM, Bidirectional LSTM, dan GRU. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(1), 671–684. <https://doi.org/10.35957/jatisi.v9i1.1588>

Khumaidi, A., Raafi`udin, R., & Solihin, I. P. (2020). Pengujian Algoritma Long Short-Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung. *Jurnal Telematika*, 15(1), 13–18.

Limbong, R., Aziz Luthfi, A. A., Yufitri, S., Chandra, A. F., & Bin Ghazali, M. (2023). Kesalehan Ekologis Masyarakat Muslim Pekanbaru: Studi Terhadap Hadis Dalam Upaya Meminimalisir Kerusakan Lingkungan. *Harmoni*, 22(1), 70–92. <https://doi.org/10.32488/harmoni.v1i22.617>

Loelianto, I., Thayf, M. S. S., & Angriani, H. (2020). Implementasi Teori Naive Bayes Dalam Klasifikasi Calon Mahasiswa Baru Stmik Kharisma Makassar. *SINTECH (Science and Information Technology) Journal*, 3(2), 110–117. <https://doi.org/10.31598/sintechjournal.v3i2.651>

Mustafa, Sunuh, H. S., Subagyo, I., & Bungawati, A. (2021). Pencemaran Udara dan ISPA. In *Angewandte Chemie International Edition*, 6(11), 951–952.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengummumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Nababan, A. A., Jannah, M., Aulina, M., & Andrian, D. (2023). Prediksi Kualitas Udara Menggunakan Xgboost Dengan Synthetic Minority Oversampling Technique (Smote) Berdasarkan Indeks Standar Pencemaran Udara (Ispu). *JTIK (Jurnal Teknik Informatika Kaputama)*, 7(1), 214–219. <https://doi.org/10.59697/jtik.v7i1.66>
- Novelan, M. S. (2020). Sistem Monitoring Kualitas Udara Dalam Ruang Menggunakan Mikrokontroler dan Aplikasi Android. *InfoTekJar :Jurnal Nasional Informatika Dan Teknologi Jaringan*, 4(2), 50–54. <https://doi.org/10.30743/infotekjar.v4i2.2306>
- Oktaviani, A., & Hustinawati. (2021). Prediksi Rata-Rata Zat Berbahaya Di Dki Jakarta Berdasarkan Indeks Standar Pencemar Udara Menggunakan Metode Long Short-Term Memory. *Jurnal Ilmiah Informatika Komputer*, 26(1), 41–55. <https://doi.org/10.35760/ik.2021.v26i1.3702>
- Prabowo, H. (2019). *Surabaya Untuk Menentukan Kategori*.
- Pradana, R., & Astika, I. B. P. (2019). Pengaruh Ukuran Perusahaan, Penerapan Good Corporate Governance, dan Pengungkapan Corporate Social Responsibility pada Nilai Perusahaan. *E-Jurnal Akuntansi*, 28(3), 1920. <https://doi.org/10.24843/eja.2019.v28.i03.p18>
- Rahayu, P., Sudipa, I. G. I., Suryani, Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M., Slamet, I., Harlina, S., & May Sanjaya, I. M. (2018). *Buku Ajar Data Mining* (Vol. 1, Issue January 2024).
- RI, P. P. (1999). *Peraturan Pemerintah Republik Indonesia nomor 41 tahun 1999 Tentang Parameter Pencemaran Udara*.
- RI, P. P. (2021). *Peraturan Pemerintah No. 22 tahun 2021 Tentang Baku Mutu Ambien Nasional*.
- Riska Chairunisa, Adiwijaya, & Widi Astuti. (2020). Perbandingan CART dan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Random Forest untuk Deteksi Kanker berbasis Klasifikasi Data Microarray. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 805–812. <https://doi.org/10.29207/resti.v4i5.2083>

Rofiq, H., Pelangi, K. C., & Lasena, Y. (2020). Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naive Bayes. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 3(1), 8–15. <http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/19417.pdf>

Santoso, I., Juanda, & Dahlan, A. (2022). *Pengawasan Kualitas Udara Dalam Gedung*. 1–23.

Siringoringo, R., Perangin Angin, R., & Rumahorbo, B. (2022). Model Klasifikasi Genetic-XGBoost Dengan T-Distributed Stochastic Neighbor Embedding Pada Peramalan Pasar. *Jurnal Times*, XI(1), 30–36. <https://archive.ics.uci.edu/ml/datasets/online+retail>

Subagiyo, H., Tri Wahyuni, R., Akbar, M., & Ulfa, F. (2021). Rancang Bangun Sensor Node untuk Pemantauan Kualitas Udara. *Jurnal Sains, Teknologi Dan Industri*, 18(1), 72. <https://doi.org/10.24014/sitekin.v18i1.11461>

Sundari, L. (2020). *Pemodelan Time Series Untuk Peramalan Suhu Udara Menggunakan Metode Long Short Term Memory (LSTM) (Studi Kasus: Stasiun Klimatologi Lampung)*. July, 1–23. July, 1–23.

Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika*. *Jurnal Edik Informatika*, 2(2), 213–219.

Yusuf, A., Kusriani, K., & Muhammad, A. H. (2022). Perbandingan Additive dan Multiplicative Exponential Smoothing Terhadap Prakiraan Kualitas Udara di Banjarmasin. *Jurnal ELTIKOM*, 6(1), 40–55. <https://doi.org/10.31961/eltikom.v6i1.507>



Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR RIWAYAT HIDUP

Data Pribadi

Nama Lengkap : Rezki Quratul. NF
Tempat/Tanggal Lahir : Sasak, 10 April 2001
Jenis Kelamin : Laki- laki
Agama : Islam
Kewarganegaraan : Indonesia
Alamat : Jorong Pd. Halaban, Nagari Sasak Ranah Pasisie, Kec. Sasak Ranah Pasisie, kab. Pasaman barat
No. Hp : 0823 8513 0794
Email : 12050110320@students.uin-suska.ac.id



Pendidikan

2006 – 2007 : TK 'Aisyah Bustanul Athfal Sasak Ranah Pasisie
2007 – 2014 : SD N 01 Sasak Ranah Pasisie
2014 – 2017 : MTS N 04 Pasaman barat
2017 – 2020 : SMK Ibnu Taimiyah Pekanbaru
2020 – 2024 : S1 Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan syarif Kasim Riau

- Hak Cipta Diindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.