PREDIKSI CURAH HUJAN MENGGUNAKAN METODE ARIMA (*AUTOREGRESSIVE INTEGRATED MOVING AVERAGE*) (STUDI KASUS : KABUPATEN SLEMAN, SEMARANG, DAN SURABAYA)



Disusun Oleh:

Nama : Hafidh Adiyatma Ramadhan
NIM : 20523039

PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
2025

HALAMAN PENGESAHAN DOSEN PEMBIMBING

PREDIKSI CURAH HUJAN MENGGUNAKAN METODE ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE) (STUDI KASUS: KABUPATEN SLEMAN, SEMARANG, DAN SURABAYA)

TUGAS AKHIR

Disusun Oleh:

Nama : Hafidh Adiyatma Ramadhan
NIM : 20523039

Yogyakarta, 1 Januari 2025

Pembimbing,

25/02/2025

(Irving Vitra Paputungan, S.T., M.Sc., Ph.D.)

HALAMAN PENGESAHAN DOSEN PENGUJI

PREDIKSI CURAH HUJAN MENGGUNAKAN METODE ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE) (STUDI KASUS: KABUPATEN SLEMAN, SEMARANG, DAN SURABAYA)

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 24 Januari 2025

Tim Penguji

Irving Vitra Paputungan, S.T., M.Sc., Ph.D. — 25/02/2025

Anggota 1

Dr. Nur Wijayaning Rahayu, S.Kom., M.Cs.

Anggota 2

Sri Mulyati, S.Kom., M.Kom.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia

(Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama:Hafidh Adiyatma Ramadhan

NIM:20523039

Tugas akhir dengan judul:

PREDIKSI CURAH HUJAN MENGGUNAKAN METODE ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE) (STUDI KASUS: KABUPATEN SLEMAN, SEMARANG, DAN SURABAYA)

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 24 Januari 2025



(Hafidh Adiyatma Ramadhan)

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur ke hadirat Allah SWT atas segala nikmat kesehatan, kekuatan, dan kesempatan, penulis mempersembahkan tugas akhir ini kepada kedua orang tua tercinta, atas kasih sayang, dukungan, dan doa yang tulus di setiap langkah hidup penulis. Penulis menyampaikan ucapan terima kasih kepada Bapak Irving Vitra Paputungan, S.T., M.Sc., Ph.D., selaku dosen pembimbing, dan Bapak Rahadian Kurniawan, S.Kom., M.Kom., selaku dosen pembimbing akademik, yang telah memberikan arahan kepada penulis dengan penuh kesabaran hingga tugas akhir ini selesai. Serta kepada seluruh dosen dan karyawan di Program Studi Informatika Universitas Islam Indonesia, yang telah mendukung perjalanan akademik penulis. Penulis juga mengucapkan apresiasi kepada Raisa Salsabila atas dukungan, motivasi, dan semangat yang tak pernah lelah diberikan kepada penulis, serta teman-teman Informatika angkatan 2020, terutama Alief, Haifan, dan Seno, atas dukungan dan kebersamaan yang tak ternilai. Semoga karya ini bermanfaat dan dapat memberi kontribusi positif bagi para pembaca.

HALAMAN MOTO

Man Jadda Wajada

“If you never try, you never know” - Coldplay

KATA PENGANTAR

Assamu'alaikum Warahmatullahi Wabarakatuh,

Alhamdulillahirabbil'alamin, puji, syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan judul "Prediksi Curah Hujan Menggunakan Metode Arima (Autoregressive Integrated Moving Average) (Studi Kasus: Kabupaten Sleman, Semarang, Dan Surabaya)". Tugas akhir ini disusun sebagai salah satu syarat untuk menyelesaikan Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Islam Indonesia.

Penulis ingin menyampaikan terima kasih yang sebesar - besarnya kepada pihak – pihak yang telah terlibat dalam memberi arahan, masukan, bimbingan, serta dukungan kepada penulis, sehingga penulis dapat menyelesaikan tugas akhir ini.

Ucapan terima kasih khusus penulis sampaikan kepada:

1. Allah SWT yang telah memberikan nikmat Kesehatan, kekuatan, dan kesempatan sehingga penulis dapat menyelesaikan tugas akhir hingga selesai.
2. Kedua orang tua penulis, Bapak Agus Wiryadi dan Ibu Eko Prapti Handayani, yang telah memberikan semangat yang tidak pernah berhenti untuk penulis Terima kasih atas segala kasih sayang, pengorbanan, dukungan, motivasi, serta doa yang tulus dalam setiap langkah yang penulis tempuh.
3. Taufiq Akbar Pratama sebagai kakak penulis yang telah memberikan semangat kepada penulis dalam penyusunan tugas akhir ini.
4. Bapak Irving Vitra Paputungan, S.T., M.Sc., Ph.D. selaku dosen pembimbing tugas akhir penulis, yang telah membimbing penulis dalam penyusunan tugas akhir ini hingga selesai.
5. Bapak Rahadian Kurniawan, S.Kom., M.Kom. selaku dosen pembimbing akademik penulis, yang telah membimbing penulis dari awal masa perkuliahan hingga masa akhir perkuliahan.
6. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc. selaku Ketua Jurusan Informatika Universitas Islam Indonesia.
7. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku Ketua Program Studi Informatika Program Sarjana.

8. Raisa Salsabila yang tidak pernah lelah untuk memberikan motivasi, semangat, serta dukungan kepada penulis untuk menyelesaikan tugas akhir ini dan juga selalu menghibur penulis dikala penulis sedang suntuk.
9. Teman – teman jurusan informatika angkatan 2020 terutama Alief, Haifan, Seno dan Rifaldi yang telah memberikan semangat, motivasi kepada penulis dalam penyusunan tugas akhir ini.

Penulis menyadari laporan ini masih memiliki kekurangan dan terbuka untuk saran serta kritik yang membangun demi perbaikan. Semoga tugas akhir ini memberi manfaat dan kontribusi positif bagi para pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Yogyakarta, 24 Januari 2025



(Hafidh Adiyatma Ramadhan)

SARI

Di tengah perubahan iklim yang semakin kompleks, pemahaman yang mendalam tentang pola curah hujan menjadi sangat penting untuk menghadapi dampak cuaca ekstrem. Penelitian ini bertujuan untuk menerapkan metode *Autoregressive Integrated Moving Average* (ARIMA) dalam memprediksi curah hujan di tiga wilayah berbeda di Indonesia, yaitu Kabupaten Sleman, Semarang, dan Surabaya. Proses penelitian mencakup pengumpulan data curah hujan dari stasiun BMKG untuk periode 2020 hingga 2022, pembersihan dan pengolahan data, serta validasi model menggunakan *Mean Squared Error* (MSE) dan *Mean Absolute Percentage Error* (MAPE). Hasil analisis menunjukkan bahwa model ARIMA (0, 1, 1) paling sesuai untuk Kabupaten Sleman dengan nilai MSE sebesar 0.865332 dan MAPE sebesar 8.53%. Sementara model ARIMA (1, 1, 1) lebih efektif untuk Kota Semarang dan Surabaya dengan MSE sebesar 0.0911438 dan MAPE sebesar 9.11% untuk Kota Semarang dan MSE sebesar 0.157882 dan MAPE sebesar 15.78% untuk Surabaya.

Kata kunci: ARIMA, Cuaca, MAPE, MSE, Prediksi

GLOSARIUM

ACF	mengukur korelasi antara nilai data pada waktu tertentu dengan nilai data pada waktu sebelumnya dalam deret waktu.
ARIMA	model deret waktu untuk memprediksi data historis menggunakan komponen autoregressive, differencing, dan moving average.
Differencing	teknik menghilangkan tren dalam data agar data menjadi stasioner.
Forecasting	proses memprediksi nilai masa depan berdasarkan data historis.
MAPE	metrik yang menunjukkan rata-rata persentase kesalahan dalam prediksi model.
MSE	metrik yang menunjukkan rata-rata kuadrat kesalahan prediksi model.
PACF	mengukur korelasi antara nilai data pada waktu tertentu dengan nilai data sebelumnya, dengan menghilangkan pengaruh nilai antaranya.
Stasioner	tidak berubah.
Time Series	jenis data berdasarkan waktu.

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PENGESAHAN DOSEN PEMBIMBING ii
HALAMAN PENGESAHAN DOSEN PENGUJI iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR iv
HALAMAN PERSEMBAHAN v
HALAMAN MOTO vi
KATA PENGANTAR vii
SARI ix
GLOSARIUM x
DAFTAR ISI xi
DAFTAR TABEL xiii
DAFTAR GAMBAR xiv
BAB I PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah dan Pertanyaan Penelitian 3
1.3 Batasan Penelitian 3
1.4 Tujuan Penelitian 3
1.5 Manfaat Penelitian 4
1.6 Metodologi Secara Umum 4
1.7 Struktur Laporan 5
BAB II KAJIAN PUSTAKA 6
2.1 Curah Hujan 6
2.2 Deret Waktu 6
2.3 Teknik dalam Memprediksi Curah Hujan 7
2.3.1 Pendekatan Subjektif 7
2.3.2 Pendekatan Objektif 8
BAB III METODOLOGI PENELITIAN 18
3.1 Tahapan ARIMA 18
3.2 Pengambilan Data 19
3.3 Processing Data 21
3.4 Analisis Awal Data 23
3.4.1 Plot Time Series 24

3.4.2 Uji Stasioneritas 24
3.4.3 Plot ACF dan PACF 25
3.4.4 Differencing 26
3.5 Validasi Model 26
3.6 Evaluasi Model 27
BAB IV HASIL DAN PEMBAHASAN 28
4.1 Detail Spesifikasi Perangkat Keras dan Lunak yang Digunakan 28
4.1.1 Perangkat Keras (*Hardware*) 28
4.1.2 Perangkat Lunak (*Software*) 28
4.2 Analisis Awal Data 28
4.2.1 Plot Time Series 29
4.2.2 Uji Stasioneritas 31
4.2.3 Plot ACF dan PACF 35
4.2.4 Differencing 39
4.3 Validasi Model 41
4.4 Evaluasi Model 43
4.4.1 *Mean Squared Error* (MSE) 43
4.4.2 *Mean Absolute Percentage Error* (MAPE) 44
4.4.3 Tabel Range MSE dan MAPE 45
4.4.4 Kesimpulan Model yang Cocok 45
4.5 Peramalan Curah Hujan 46
4.5.1 Kesimpulan Peramalan Curah Hujan 49
BAB V KESIMPULAN DAN SARAN 50
5.1 Kesimpulan 50
5.2 Saran 50
DAFTAR PUSTAKA 51
LAMPIRAN 54

DAFTAR TABEL

Tabel 2.1 Tabel Perbandingan Pendekatan Objektif.....15
Tabel 3.1 Contoh data curah hujan wilayah Kabupaten Sleman 20
Tabel 3.2 Contoh data curah hujan wilayah Kota Semarang.....20
Tabel 3.3 Contoh data curah hujan wilayah Surabaya.....21
Tabel 3.4 Contoh pemeriksaan data awal 21
Tabel 3.5 Contoh penggantian nilai kosong dengan median 22
Tabel 3.6 Contoh penggantian nilai setelah mendapatkan median.....23
Tabel 4.1 Ringkasan Data Curah Hujan 29
Tabel 4.2 Model ARIMA Kabupaten Sleman 41
Tabel 4.3 Model ARIMA Semarang.....42
Tabel 4.4 Model ARIMA Surabaya.....43
Tabel 4.5 MSE Wilayah Penelitian.....44
Tabel 4.6 MAPE Wilayah Penelitian.....44
Tabel 4.7 Range MSE dan MAPE 45
Tabel 4.8 Evaluasi MSE dan MAPE.....45
Tabel 4.9 Kategori Curah Hujan Bulanan.....49

DAFTAR GAMBAR

Gambar 2.1 Jaringan Syaraf Tiruan (JST) Multilayer Perceptron 10
Gambar 2.2 Arsitektur LSTM (Long Short-Term Memory) 12
Gambar 2.3 Struktur Decision Tree 13
Gambar 2.4 Ilustrasi Random Forest 14
Gambar 3.1 Tahapan ARIMA 18
Gambar 3.2 Contoh Plot Time Series Kabupaten Sleman 24
Gambar 3.3 Contoh Uji Stasioneritas Kabupaten Sleman 24
Gambar 3.4 Contoh ACF dan PACF Kabupaten Sleman 26
Gambar 3.5 Contoh Differencing Kabupaten Sleman 26
Gambar 4.1 Plot Time Series Kabupaten Sleman 29
Gambar 4.2 Plot Time Series Semarang 30
Gambar 4.3 Plot Time Series Surabaya 30
Gambar 4.4 Nilai Uji Stasioneritas Kabupaten Sleman Sebelum Transformasi 31
Gambar 4.5 Nilai Uji Stasioneritas Kabupaten Sleman Setelah Transformasi 31
Gambar 4.6 Nilai Uji Stasioneritas Semarang Sebelum Transformasi 32
Gambar 4.7 Nilai Uji Stasioneritas Semarang Setelah Transformasi Satu Kali 33
Gambar 4.8 Nilai Uji Stasioneritas Semarang Setelah Transformasi Dua Kali 33
Gambar 4.9 Nilai Uji Stasioneritas Surabaya Sebelum Transformasi 34
Gambar 4.10 Nilai Uji Stasioneritas Surabaya Setelah Transformasi 34
Gambar 4.11 ACF dan PACF Kabupaten Sleman 35
Gambar 4.12 ACF dan PACF Semarang 36
Gambar 4.13 ACF dan PACF Surabaya 38
Gambar 4.14 Differencing Kabupaten Sleman 39
Gambar 4.15 Differencing Semarang 40
Gambar 4.16 Differencing Surabaya 41
Gambar 4.17 Hasil Plot Time Series Kabupaten Sleman ARIMA (0, 1, 1) 46
Gambar 4.18 Hasil Plot Time Series Semarang ARIMA (1, 1, 1) 47
Gambar 4.19 Hasil Plot Time Series Surabaya ARIMA (1, 1, 1) 48

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Cuaca serta iklim memiliki keterkaitan yang sangat kuat, dan pemahaman keduanya menjadi suatu hal yang krusial (Fadlan et al., 2022) dalam mengelola di berbagai sektor, terutama sektor - sektor yang dipengaruhi oleh kondisi alam. Jika prediksi cuaca tidak dilakukan secara tepat, maka berbagai dampak negatif dapat timbul, seperti kerusakan pada sektor pertanian, bencana alam seperti banjir dan tanah longsor, serta kegagalan infrastruktur (Polawan & Alam, 2019). Selain itu, kurangnya dari prediksi cuaca yang tidak akurat dapat mempersulit pengelolaan sumber daya air, yang mengakibatkan pada terjadinya penurunan kualitas hidup dan keberlangsungan ekosistem (Herlina & Prasetyorini, 2020). Secara umum cuaca merupakan kejadian dimana kondisi udara di suatu area dengan durasi waktu yang singkat. Iklim adalah kondisi rata – rata cuaca selama satu tahun yang dikaji dalam durasi waktu yang panjang sekurang - kurangnya 30 tahun serta mencakup area yang lebar (Anshari et al., 2013). Secara umum terdapat unsur – unsur yang dapat mempengaruhi suatu keadaan cuaca serta iklim di area tertentu, yaitu : tekanan udara, kelembaban udara suhu udara serta curah hujan (Safar et al., 2019). Pentingnya pemahaman yang mendalam tentang pola curah hujan tidak dapat disangkal terutama dalam mendukung sektor pertanian yang bergantung pada kondisi cuaca yang stabil (Masum, 2019). Pemahaman yang tepat juga sangat penting dalam pengelolaan sumber daya air yang semakin kompleks (Alexander & Harahab, 2009), di mana pemahaman yang mendalam tentang curah hujan sangat penting untuk menjaga keberlanjutan ekosistem air dan infrastruktur.

Tidak hanya di berbagai wilayah di Indonesia yang bergantung pada sektor pertanian, namun pentingnya pemahaman akan pola curah curah hujan juga berlaku lebih luas secara global, dengan berbagai wilayah yang memiliki karakteristik geografis, ekonomi, dan budaya yang berbeda. Pemahaman yang cermat terhadap pola curah hujan menjadi landasan utama untuk mengelola risiko dari dampak terjadinya bencana yang disebabkan oleh perubahan cuaca dan meningkatkan produktivitas sektor pertanian, infrastruktur, serta bidang lainnya yang menjadi tulang punggung ekonomi di daerah-daerah tersebut. Dalam konteks ini, keberhasilan dalam meramalkan serta memahami perubahan musim curah hujan memiliki implikasi besar

terhadap strategi pengelolaan sumber daya air yang semakin kompleks (Kim et al., 2007). Curah hujan yang tidak terduga dapat mengakibatkan dampak yang signifikan pada pertanian, infrastruktur, bencana alam seperti longsor, banjir, dan juga kekeringan (Polawan & Alam, 2019) serta keberlanjutan lingkungan di wilayah-wilayah ini, menjadikan pemahaman pola curah hujan sebagai elemen kunci dalam upaya mitigasi dan adaptasi terhadap perubahan iklim yang terus berlangsung (Aulia et al., 2018). Ketersediaan informasi cuaca yang dapat diandalkan memainkan peran penting dalam memberikan peringatan dini terhadap bencana seperti banjir dan longsor yang dapat menyebabkan kerugian besar, baik secara materiil maupun non-materiil. Dengan prediksi yang kurang akurat, risiko banjir meningkat, perencanaan pertanian terhambat, dan evakuasi saat bencana alam terjadi menjadi lebih sulit (Burrahman et al., 2018). Dalam upaya memahami kondisi cuaca di wilayah penelitian yang dipilih, keputusan untuk menentukan wilayah - wilayah tersebut didasarkan pada kekhasan kondisi geografisnya. Ini tidak berarti bahwa hanya wilayah - wilayah ini yang menghadapi masalah prediksi cuaca, karena beberapa daerah lain juga mengalami tantangan serupa terkait manajemen sumber daya, pertanian, dan risiko bencana. Namun, pilihan ini dibuat dengan mempertimbangkan representasi geografis yang bervariasi dan penting, guna membantu memahami variasi cuaca yang mungkin terjadi di wilayah-wilayah berbeda. Kawasan yang dipilih mencakup beragam karakteristik geografis, seperti perbedaan topografi dan variasi iklim, yang memberikan gambaran lebih luas tentang pola cuaca. Meskipun daerah-daerah ini tidak sepenuhnya mewakili keseluruhan provinsi, karakteristik uniknya memberikan dasar yang kuat untuk memahami variasi cuaca di masing-masing wilayah.

Untuk mengatasi tantangan memprediksi cuaca di wilayah - wilayah yang diteliti, pendekatan objektif menggunakan metode ARIMA. Penelitian ini akan membahas bagaimana penerapan model *Autoregressive Integrated Moving Average* (ARIMA) dapat digunakan dalam memprediksi curah hujan di wilayah - wilayah tersebut. Dengan demikian, penelitian ini bertujuan menerapkan metode ARIMA untuk memprediksi curah hujan di wilayah - wilayah yang diteliti selama periode tertentu. Model ARIMA adalah sebuah metode yang diutarakan pada sekitar tahun 1970 oleh George Box dan Gwilym Jenkins (Nur & Astuti, 2006). *Autoregressive Integrated Moving Average* (ARIMA) memiliki keunggulan dalam fleksibilitasnya, mampu menyesuaikan diri dengan pola data yang ada (Shumway & Stoffer, 2000). Kemampuan ini menjadikannya pilihan yang tepat untuk meramalkan sejumlah variabel dengan cepat, secara sederhana, dan ekonomis (Hutasuhut, 2014). Dengan pemahaman yang

lebih baik tentang pola curah hujan di Kabupaten Sleman, Semarang, dan Surabaya, dapat meningkatkan kesiapsiagaan dalam menghadapi tantangan cuaca ekstrem dan mengoptimalkan pengelolaan sumber daya air. Dengan kata lain, penelitian ini bertujuan untuk memberikan wawasan berharga yang dapat digunakan oleh para pemangku kepentingan, peneliti, dan pengambil keputusan dalam menghadapi kompleksitas tantangan yang dihadapi oleh wilayah-wilayah ini dalam konteks iklim yang berubah dengan cepat (Aulia et al., 2018).

1.2 Rumusan Masalah dan Pertanyaan Penelitian

Penyebutan Berdasarkan latar belakang masalah di atas, rumusan masalah yang akan saya bahas pada penelitian ini adalah:

a. Bagaimana penerapan model *Autoregressive Integrated Moving Average* (ARIMA) dapat digunakan dalam memprediksi curah hujan di Kabupaten Sleman, Semarang, dan Surabaya.
b. Apa keunggulan metode *Autoregressive Integrated Moving Average* (ARIMA) dibandingkan dengan pendekatan objektif lainnya dalam hal efisiensi dan penerapannya?

Pertanyaan penelitian yang diangkat adalah:

a. Apakah terdapat perbedaan signifikan dalam kinerja model ARIMA antara ketiga wilayah?

1.3 Batasan Penelitian

Batasan masalah yang ditentukan berdasarkan rumusan masalah penelitian ini yaitu:

a. Data yang digunakan dalam penelitian ini adalah data cuaca historis yang tersedia dan relevan, serta data curah hujan yang diamati dalam periode waktu tertentu.
b. Data yang digunakan dalam penelitian ini didapatkan dari website BMKG (Badan Meteorologi, Klimatologi dan Geofisika).

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah menerapkan metode *Autoregressive Integrated Moving Average* (ARIMA) untuk memprediksi curah hujan di Kabupaten Sleman, Semarang, dan Surabaya selama periode waktu tertentu.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

a. Hasil dari penelitian ini diharapkan mampu memperkaya referensi serta literatur khususnya tentang prediksi curah hujan menggunakan metode ARIMA (*Autoregressive Integrated Moving Average*).
b. Penelitian ini diharapkan dapat memperluas pemahaman tentang ARIMA digunakan untuk memprediksi curah hujan.

1.6 Metodologi Secara Umum

Penelitian ini menggunakan pendekatan objektif dengan menggunakan metode ARIMA (*Autoregressive Integrated Moving Average*) dalam memprediksi curah hujan di wilayah Kabupaten Sleman, Semarang, dan Surabaya. Untuk menerapkan metode ini, penelitian dilakukan melalui beberapa tahapan sebagai berikut:

a. Sebagai tahap awal, data curah hujan harian dikumpulkan dari stasiun meteorologi BMKG di masing - masing wilayah penelitian selama periode tahun 2020 hingga 2022. Data ini menjadi dasar untuk melakukan analisis lebih lanjut.
b. Sebelum dilakukan analisis, data yang telah dikumpulkan perlu melalui tahap pengolahan data. Tahap ini meliputi pembersihan data untuk menghilangkan data yang tidak relevan, imputasi data untuk mengatasi data yang hilang, serta transformasi data agar memenuhi asumsi model yang akan diperlukan.
c. Setelah data diolah, langkah selanjutnya adalah melakukan analisis data. Analisis ini bertujuan untuk mengidentifikasi pola dan karakteristik data curah hujan serta membangun model prediksi yang akurat. Untuk memvisualisasikan pola dan trend data curah hujan, dilakukan plot time series. Plot ini memberikan gambaran awal mengenai karakteristik data yang akan dimodelkan. Setelah visualisasi, dilakukan uji stasioneritas untuk memastikan bahwa data memenuhi salah satu asumsi penting dalam model ARIMA. Lalu data yang sudah melewati tahap uji stasioneritas, dilakukan identifikasi model ARIMA yang paling sesuai dengan data. Identifikasi ini dilakukan dengan cara menganalisis fungsi autokorelasi (ACF) dan fungsi autokorelasi parsial (PACF). Tahap selanjutnya adalah differencing, tahap ini berguna untuk menampilkan data yang sudah stasioner dengan menghilangkan trend atau pola data curah hujan.

d. Untuk mengukur kinerja model ARIMA yang telah dibangun, dilakukan validasi model menggunakan data yang sudah disiapkan sebelumnya. Validasi ini bertujuan untuk melihat seberapa baik model dapat memprediksi data.
e. Sebagai langkah terakhir, dilakukan evaluasi model dengan menghitung nilai MSE dan MAPE. Nilai - nilai ini digunakan untuk membandingkan kinerja berbagai model yang telah dibangun.

1.7 Struktur Laporan

Struktur penulisan laporan penelitian ini adalah sebagai berikut:

a. Bab 1 Pendahuluan

Pada bab ini membahas tentang latar belakang penelitian, rumusan dan pertanyaan penelitian, batasan penelitian, tujuan dan manfaat penelitian, metodologi penelitian secara umum, serta struktur laporan.

b. Bab 2 Kajian Pustaka

Bab ini membahas tentang kajian pustaka yang terdapat di penelitian ini.

c. Bab 3 Metodologi Penelitian

Pada bab 3 membahas tentang bagaimana tahapan ARIMA, serta bagaimana langkah langkah pengolahan data agar dapat digunakan untuk melakukan prediksi curah hujan menggunakan metode ARIMA.

d. Bab 4 Hasil Dan Pembahasan

Bab 4 ini membahas tentang hasil yang diperoleh dari penelitian yang sudah dilakukan menggunakan metode ARIMA.

e. Bab 5 Kesimpulan Dan Saran

Bab ini berisikan tentang kesimpulan dari penelitian yang telah dilakukan serta saran dari penulis.

BAB II
KAJIAN PUSTAKA

2.1 Curah Hujan

Curah hujan merujuk pada jumlah air yang turun ke permukaan tanah datar selama suatu periode tertentu, yang diukur dalam satuan tinggi (millimeter) di atas permukaan horizontal, tanpa mempertimbangkan penguapan, aliran permukaan, atau infiltrasi. Pengukuran curah hujan ini melibatkan penentuan volume air yang jatuh ke permukaan bidang datar selama rentang waktu tertentu, seperti harian, mingguan, bulanan, atau tahunan. Hujan ekstrem, yang ditandai dengan intensitas curah hujan yang tinggi, memiliki potensi untuk menyebabkan banjir (Laia & Setyawan, 2020). Rainfall rate (RR) atau laju curah hujan adalah pengukuran intensitas hujan dalam satuan milimeter per jam (mm/jam). RR digunakan untuk mengukur kecepatan jatuhnya hujan ke permukaan tanah selama periode waktu tertentu dan sering kali digunakan dalam prediksi cuaca serta model hidrologi untuk memahami potensi dampak hujan pada banjir dan aliran air permukaan.

Memprediksi curah hujan merupakan aspek yang paling menantang dan terus-menerus diperbarui dalam pemantauan cuaca. Meskipun telah terjadi peningkatan akurasi dalam memprediksi parameter cuaca seperti suhu, tekanan, dan angin, peningkatan ini belum merata dalam prediksi curah hujan. Hal ini disebabkan oleh kompleksitas proses pembentukan awan dan hujan, yang dipengaruhi oleh berbagai faktor seperti suhu udara, kelembaban,tekanan udara serta topografi wilayah Dalam memahami curah hujan, penting untuk mempertimbangkan pola dan trend yang terkait dengan fenomena ini. Salah satu cara untuk menganalisis pola curah hujan adalah dengan menggunakan deret waktu (*Times Series*). Melalui analisis deret waktu (*Time Series*) pola musiman,trend jangka panjang serta peristiwa dalam curah hujan dapat diidentifikasi secara lebih akurat.

2.2 Deret Waktu

Data deret waktu (*Time Series*) adalah serangkaian pengamatan yang tersusun berdasarkan urutan waktu dengan interval yang sama. Jenis data ini sering ditemui dalam kehidupan sehari-hari karena pengumpulan data dilakukan secara berkala, seperti harian, mingguan, atau bulanan. Dalam kumpulan data ini, terdapat pola yang dapat diamati. Dalam analisis deret waktu, pola ini biasanya diklasifikasikan ke dalam tiga kategori utama: pola tren,

siklus, dan musiman (*seasonal*). Pola musiman adalah jenis pola yang mengulang secara berkala pada interval tertentu (Al'afi et al., 2020). Dalam menganalisis data deret waktu curah hujan, pola musiman curah hujan dapat diidentifikasi. Namun, untuk membuat prediksi yang lebih akurat tentang curah hujan, diperlukan teknik yang lebih canggih.

2.3 Teknik dalam Memprediksi Curah Hujan

Dalam memprediksi curah hujan terdapat 2 pendekatan, yaitu pendekatan subjektif dan pendekatan objektif. Berikut adalah penjelasannya.

2.3.1Pendekatan Subjektif

Pendekatan subjektif adalah pendekatan dalam memprediksi curah hujan yang melibatkan penilaian pribadi ahli meteorologi, yang menggunakan pengalaman, pengetahuan, dan intuisi untuk membuat prediksi curah hujan (Murphy & Brown, 1984). Dalam pendekatan subjektif, terdapat 2 cara, yaitu:

a. Klimatologi:

Cara klimatologi adalah cara untuk memprediksi pola curah hujan berdasarkan data historis. Cara ini memiliki keunggulan dapat memprediksi pola curah hujan musiman berdasarkan dari data historis yang dimiliki selama beberapa dekade, untuk mengidentifikasi pola yang konsisten dan membuat prediksi curah hujan di masa yang akan datang, serta kelemahan dalam memprediksi pola curah hujan harian, mingguan hingga musiman dengan tingkat akurasi tinggi, dan bergantung pada ketersediaan catatan data historis yang dimiliki.

b. Heuristik Cuaca Lokal:

Cara ini mengandalkan pengetahuan lokal dan pengalaman prakiraan cuaca dari para ahli atau petani yang telah lama mengamati pola cuaca di wilayah tertentu. Heuristik cuaca lokal mempunyai keunggulan dalam memprediksi curah hujan karena cara ini menggunakan pengetahuan lokal serta pengalaman yang dikumpulkan dari para ahli atau petani yang telah lama mengamati kondisi cuaca di suatu daerah tertentu, untuk dapat membuat prediksi curah hujan yang akurat. Namun cara ini memiliki kelemahan, karena cara ini masih mengandalkan sebagian besar pada pengetahuan dan individu. dengan demikian, jika individu tidak memiliki pengetahuan atau pengalaman yang cukup, maka prediksi tidak akan akurat dan dipengaruhi faktor subjektif seperti asumsi individu.

2.3.2Pendekatan Objektif

Pendekatan objektif adalah pendekatan dalam memprediksi curah hujan dengan menggunakan metode metode numerik atau statistik (Murphy & Brown, 1984). Dalam pendekatan objektif terdapat metode yang umum digunakan dalam memprediksi curah hujan antara lain:

a. *Autoregressive Integrated Moving Average* (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) adalah sebuah metode yang diutarakan pada sekitar tahun 1970 oleh George Box dan Gwilym Jenkins (Nur & Astuti, 2006). *Autoregressive Integrated Moving Average* (ARIMA) memiliki keunggulan dalam fleksibilitasnya, mampu menyesuaikan diri dengan pola data yang ada. Kemampuan ini dikombinasikan dengan tingkat akurasi peramalan yang signifikan, menjadikannya pilihan yang tepat untuk meramalkan sejumlah variabel dengan cepat, secara sederhana, dan ekonomis. Sebagai perpaduan antara model autoregresif dan moving average, *Autoregressive Integrated Moving Average* (ARIMA) memanfaatkan data historis untuk meramalkan data masa depan secara efisien dan ekonomis (Hutasuhut et al., 2014). Model *Autoregressive Integrated Moving Average* (ARIMA) bekerja dengan cara mengasumsikan nilai suatu variabel pada waktu tertentu dipengaruhi oleh nilai – nilai sebelumnya (autoregressive) dan dipengaruhi oleh kesalahan acak pada waktu – waktu sebelumnya (moving average).

Sebagai perpaduan dari model AR dan MA, model *Autoregressive Integrated Moving Average* (ARIMA) yang dikenal sebagai Box-Jenkins mampu menganalisis deret waktu yang menunjukkan tren, baik itu stasioner maupun non-stasioner (Shumway & Stoffer, 2000). *Autoregressive Integrated Moving Average* (ARIMA) merupakan salah satu metode yang umum digunakan dalam melakukan analisis deret waktu karena kesederhanaanya dan kemampuannya dalam memberikan hasil yang baik, bahkan menggunakan data dengan ukuran sampel yang relatif kecil.

Model Autoregressive (AR) orde p mengasumsikan bahwa nilai suatu observasi pada waktu sekarang bergantung secara linear pada p observasi sebelumnya, membentuk suatu pola ketergantungan antara data-data dalam deret waktu (Wulandari & Gernowo, 2019). Orde p menentukan kompleksitas hubungan antara observasi saat ini dan observasi sebelumnya. Model (AR) efektif dalam menangkap pola autokorelasi dalam data yang relatif stabil, namun

kurang optimal jika data yang digunakan menganduk fluktuasi yang sangat besar, karena dapat mengganggu asumsi linearitas model.

Adapun rumus dari AR (2.1):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t \quad (2.1)$$

Model Moving Average (MA) digunakan untuk menjelaskan bagaimana suatu pengamatan pada waktu t diungkapkan sebagai hasil dari gabungan linear dari sejumlah residual atau nilai – nilai kesalahan pada periode sebelumnya. Dalam bentuk persamaannya, model MA dapat direpresentasikan sebagai suatu persamaan di mana nilai pada waktu t dijelaskan oleh kombinasi linear dari sejumlah nilai residual sebelumnya (Wulandari & Gernowo, 2019).

Adapun rumus dari MA (2.2):

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad (2.2)$$

Model Autoregressive Moving Average (ARIMA) merupakan sebuah fungsi yang bergantung pada tiga parameter utama: p sebagai orde dari komponen Autoregressive (AR), d sebagai orde dari proses differencing, dan q sebagai orde dari Moving Average (MA). Data time series dianggap stasioner setelah mengalami proses differencing sebanyak d kali, yang berarti dilakukan pengurangan antara nilai pengamatan pada saat ini dengan nilai pengamatan pada periode sebelumnya sebanyak d kali, maka ARIMA dapat digunakan untuk memodelkan hubungan antara nilai – nilai dalam time series dengan memperhitungkan orde – orde AR, differencing, dan MA yang telah ditentukan (Wulandari & Gernowo, 2019).

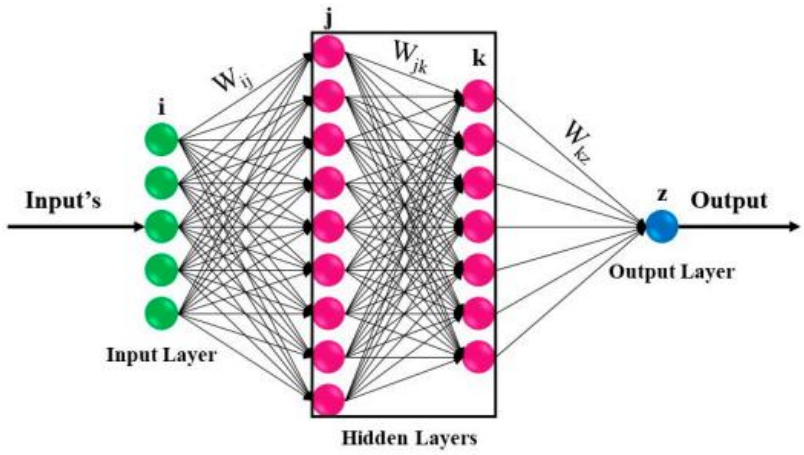
Adapun rumus dari ARIMA (2.3):

dengan $\phi_p(B)$ merupakan komponen AR orde p
dan $\theta_q(B)$ merupakan komponen orde q.

$$\phi_p(B)(1-B)^d Z_t = \mu' + \theta_q(B) a_t \quad (2.3)$$

b. Artificial Neural Network (ANN):

Artificial Neural Network atau Jaringan Syaraf Tiruan (JST) merupakan model komputasi yang mengadopsi prinsip kerja jaringan saraf biologis untuk mengolah informasi kompleks. JST terdiri dari sejumlah besar unit pemroses sederhana yang terhubung secara kompleks, memungkinkan sistem untuk mempelajari pola dalam data yang kompleks dan melakukan generalisasi terhadap data baru. Kemampuan ini menjadikan JST sebagai alat yang ampuh dalam berbagai aplikasi kecerdasan buatan, seperti pengenalan pola dan pemodelan prediksi (Watta & Hassoun, 1995). Meskipun demikian, pengembangan (JST) masih menghadapi beberapa tantangan, seperti kebutuhan akan data pelatihan dengan jumlah yang besar, pelatihan model yang membutuhkan waktu yang lama, serta interpretasi model yang rumit. Gambar 2.1 memberikan visualisasi tentang bagaimana data diproses dalam jaringan syaraf tiruan (JST).



Gambar 2.1 Jaringan Syaraf Tiruan (JST) Multilayer Perceptron

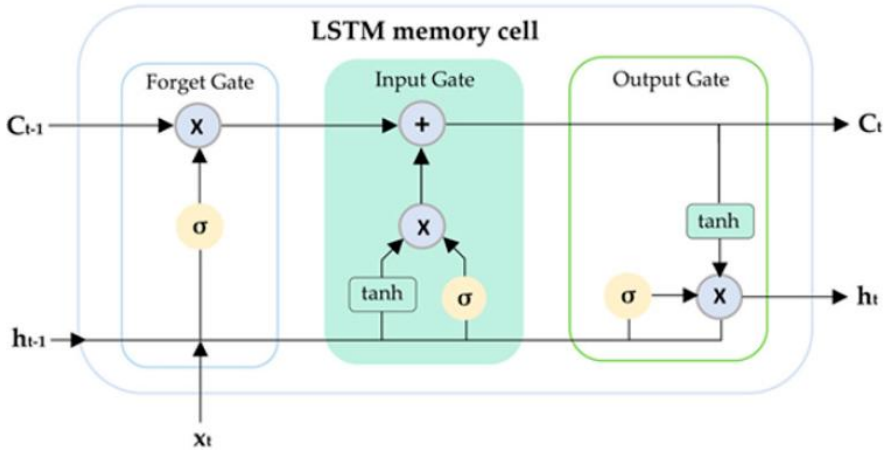
Sumber: (Haghshenas et al., 2020)

Gambar 2.1 merupakan salah satu jenis arsitektur jaringan syaraf tiruan (JST) multilayer perceptron yang paling umum digunakan. Komponen utama dalam multilayer perceptron terdiri dari tiga layer yaitu input layer, hidden layer, dan output layer. Input layer akan menerima data dan akan dilanjutkan ke layer selanjutnya yaitu hidden layer. Pada hidden layer data akan diproses melalui serangkaian perhitungan yang melibatkan bobot sinaptik dan fungsi aktivasi. Setelah selesai pada tahap hidden layer, data diteruskan menuju layer terakhir yaitu output layer yang akan menghasilkan prediksi. Setiap koneksi antara node (neuron) pada lapisan yang berbeda memiliki bobot yang dapat diubah selama proses pelatihan. Bobot ini menentukan kekuatan pengaruh suatu node (neuron) terhadap neuron lainnya. Fungsi aktivasi berperan dalam menentukan apakah suatu node (neuron) akan aktif atau tidak, sehingga

memungkinkan JST untuk mempelajari pola yang kompleks dalam data (Haghshenas et al., 2020). Keunggulan Artificial Neural Network (ANN) terletak pada kemampuannya yang fleksibel dalam mengolah suatu data yang sangat kompleks dan kelebihannya untuk belajar dari struktur data yang tidak linier. Artificial Neural Network (ANN) efektif untuk prediksi jangka panjang dan pendek. Namun, untuk hasil yang optimal Artificial Neural Network (ANN) memerlukan data pelatihan yang besar dan penyesuaian yang cermat untuk menghindari overfitting.

c. LSTM (Long Short-Term Memory):

Jaringan saraf Long Short-Term Memory (LSTM) merupakan arsitektur yang sangat efektif dalam memproses data urutan, seperti data deret waktu (time series) atau teks. Uniknya, LSTM dirancang untuk mengatasi keterbatasan jaringan saraf rekuren (RNN) tradisional dalam menangkap dependensi jangka panjang. Dengan adanya sel memori dan mekanisme gerbang yang kompleks, LSTM dapat mengingat informasi relevan dalam jangka waktu yang lama, bahkan saat dihadapkan pada data dengan ketergantungan yang jauh. Sel memori pada LSTM berperan sebagai unit penyimpanan informasi. Informasi yang masuk ke dalam sel memori akan diproses oleh tiga gerbang utama: input gate, output gate, dan forget gate (Salman et al., 2018). Input gate, di sisi lain, mengontrol aliran informasi baru yang akan disimpan dalam sel memori. Informasi baru ini akan dikombinasikan dengan informasi yang sudah ada dalam sel memori untuk menghasilkan output. Output gate menentukan bagian mana dari informasi dalam sel memori yang akan dioutputkan sebagai hasil prediksi. Forget gate berfungsi untuk menentukan bagian mana dari informasi sebelumnya yang perlu dilupakan atau dihapus dari sel memori. Hal ini penting untuk menghindari akumulasi informasi yang tidak relevan dan menjaga agar model tetap fokus pada informasi yang penting.



Gambar 2.2 Arsitektur LSTM (Long Short-Term Memory)

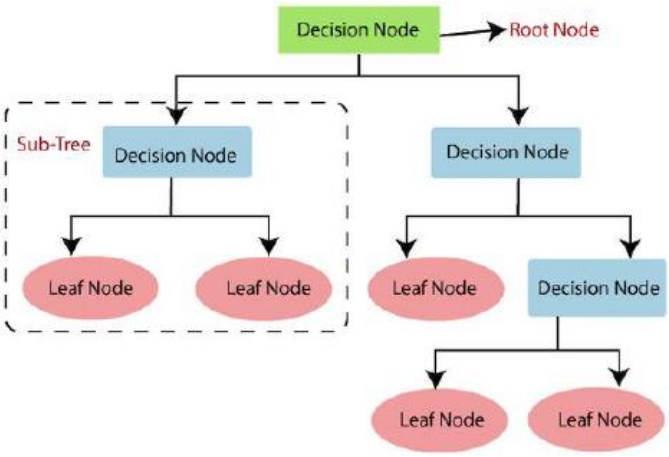
Sumber: (Nifa et al., 2023)

Gambar 2.2 menjelaskan tentang arsitektur LSTM (Long Short-Term Memory). Forget gate berfungsi untuk menentukan informasi yang harus dihapus dari keadaan sel sebelumnya (Ct-1) dengan menggunakan fungsi aktivasi sigmoid (σ), lalu hasil fungsi dikalikan dengan keadaan sel sebelumnya untuk memutuskan seberapa banyak informasi yang perlu dihilangkan. Selanjutnya input gate memiliki dua fungsi utama yaitu: fungsi sigmoid (σ) yang akan menentukan bagian dari input baru (xt) yang akan diperbarui, dan fungsi tanh yang menghasilkan nilai baru untuk ditambahkan kedalam sel. Nilai – nilai ini kemudian digabungkan untuk memperbarui keadaan sel yang baru (Ct). Lalu output gate menghasilkan nilai berdasarkan keadaan yang telah diperbarui, menggunakan fungsi sigmoid (σ) dan fungsi tanh untuk menentukan hidden state (ht) (Nifa et al., 2023). Meskipun LSTM (Long Short-Term Memory) memiliki banyak keunggulan, seperti fleksibilitas dalam menangani berbagai panjang urutan data dan kemampuannya untuk belajar representasi fitur yang kompleks, model ini juga memiliki beberapa tantangan, seperti kebutuhan akan daya komputasi yang tinggi untuk melatih model yang besar, serta kesulitan dalam menginterpretasi keputusan yang dibuat oleh model.

d. Decision Tree

Decision Tree atau dalam bahasa indonesia pohon keputusan adalah algoritma pembelajaran mesin yang digunakan untuk memprediksi hasil suatu kejadian berdasarkan serangkaian keputusan yang diambil dari atribut-atribut yang relevan. Pohon keputusan memodelkan proses pengambilan keputusan yang kompleks dengan cara yang sistematis dan mudah dipahami (Rokach & Maimon, 2006). Proses dimulai dengan node akar yang mewakili

fenomena yang ingin diprediksi. Dari node ini, atribut yang relevan adalah membagi data menjadi subset yang lebih kecil melalui node internal, yang menciptakan cabang keputusan. Keunggulan utama pohon keputusan adalah kemampuannya untuk menyajikan informasi secara intuitif, sehingga mudah diinterpretasikan oleh pengguna. Meskipun efektif, pohon keputusan juga memiliki kelemahan, seperti kecenderungan untuk overfitting jika tidak diterapkan dengan teknik pemangkasan yang tepat. Gambar 2.3 memberikan visualisasi tentang bagaimana proses decision tree.



Gambar 2.3 Struktur Decision Tree

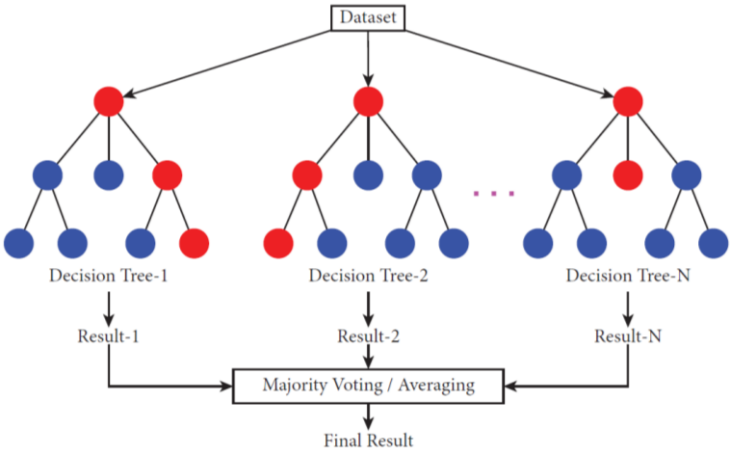
Sumber: (Fatimah & Rahmawati, 2022)

Gambar 2.3 merupakan struktur dari decision tree atau pohon keputusan. Proses kerja pohon keputusan atau decision tree dimulai dari akar pohon (root node), lalu data dibagi menjadi kelompok – kelompok yang lebih kecil berdasarkan nilai yang paling relevan (Fatimah & Rahmawati, 2022). Proses pembagian ini dilakukan berulang hingga setiap kelompok data pada daun memiliki nilai konsistensi yang tinggi, yang artinya sebagian besar data dalam kelompok tersebut memiliki kelas atau nilai yang sama sehingga proses ini akan terus berulang hingga semua data terklasifikasi membentuk daun – daun pada pohon keputusan (decision tree).

Pohon keputusan (decision tree) menawarkan kemudahan interpretasi dan fleksibilitas dalam menangani berbagai jenis data, namun memiliki keterbatasan dalam memprediksi variabel kontinu dan rentan terhadap overfitting (Rokach & Maimon, 2006). Hasil pohon keputusan (decision tree) sangat dipengaruhi oleh kualitas data yang digunakan, sehingga penting untuk melakukan preprocessing data yang baik sebelum membangun model.

e. Random Forest:

Random Forest adalah sebuah algoritma pembelajaran mesin yang sangat populer dan serbaguna. Algoritma ini bekerja berdasarkan prinsip ensemble learning, yaitu menggabungkan banyak model sederhana, dalam hal ini pohon keputusan, untuk membentuk satu model yang lebih kompleks dan akurat. Saat dilatih, random forest membangun sejumlah besar pohon keputusan, masing-masing menggunakan sampel data dan subset fitur yang berbeda secara acak. Dengan cara ini, setiap pohon akan memiliki karakteristik yang unik dan tidak terlalu berkorelasi satu sama lain. Setelah semua pohon selesai dilatih, random forest akan menggabungkan prediksi dari setiap pohon untuk menghasilkan keputusan akhir (Singh et al., 2019). Untuk tugas klasifikasi, kelas yang paling sering diprediksi oleh pohon-pohon ini akan menjadi output akhir, sedangkan untuk tugas regresi, nilai rata-rata dari semua prediksi akan digunakan. Salah satu keunggulan utama dari random forest adalah kemampuannya untuk mengurangi overfitting, sebuah masalah umum pada model pohon keputusan tunggal (Breiman, 2001). Dengan menggunakan banyak pohon yang dibangun secara acak, random forest dapat meningkatkan generalisasi model dan mengurangi varians. Selain itu, random forest juga mampu menangani dataset dengan dimensi tinggi dan mengukur pentingnya setiap fitur, sehingga membantu dalam memahami hubungan antara fitur dan target. Random Forest juga dapat diaplikasikan dalam peramalan deret waktu, di mana algoritma ini digunakan untuk memprediksi nilai di masa depan berdasarkan data historis (Tyralis & Papacharalampous, 2017). Meskipun demikian, kinerja random forest sangat bergantung pada pemilihan parameter yang tepat, seperti jumlah pohon dan jumlah fitur yang digunakan pada setiap split.



Gambar 2.4 Ilustrasi Random Forest

Sumber: (Khan et al., 2021)

Gambar 2.4 memberikan ilustrasi yang jelas mengenai cara kerja algoritma random forest. Algoritma ini bekerja dengan menggabungkan banyak pohon keputusan (decision trees) yang dibangun secara independen. Setiap pohon keputusan dilatih pada sampel data yang berbeda-beda yang diambil secara acak dari dataset keseluruhan. Dataset merupakan kumpulan data awal yang menjadi bahan baku dalam membangun model random forest. Data ini kemudian dibagi-bagi menjadi beberapa bagian yang lebih kecil untuk melatih setiap pohon keputusan. Pohon keputusan sendiri merupakan model prediksi yang sederhana, di mana setiap cabang mewakili suatu keputusan berdasarkan nilai fitur tertentu. Proses pelatihan setiap pohon keputusan melibatkan pemilihan sampel data dan fitur secara acak. Hal ini bertujuan untuk menciptakan keragaman di antara pohon-pohon tersebut, sehingga model random forest menjadi lebih kuat dan tahan terhadap overfitting (Khan et al., 2021). Setelah semua pohon selesai dilatih, masing-masing pohon akan memberikan prediksi terhadap data baru. Hasil prediksi dari setiap pohon kemudian digabungkan melalui mekanisme voting atau averaging. Jika masalah yang dihadapi adalah masalah klasifikasi, maka akan dilakukan voting untuk memilih kelas yang paling banyak dipilih oleh pohon-pohon. Sedangkan jika masalahnya adalah masalah regresi, maka akan dilakukan perhitungan rata-rata dari semua prediksi. Hasil akhir dari proses voting atau averaging ini merupakan prediksi akhir dari model random forest.

Secara umum, semua metode yang terdapat pada pendekatan objektif dapat dibandingkan seperti dalam tabel berikut.

Tabel 2.1 Tabel Perbandingan Pendekatan Objektif

Metode	**Keunggulan**	**Kekurangan**	**Penggunaan Umum**
ARIMA (*Autoregressive Integrated Moving Average)*	● Fleksibel dengan pola data ● Mempunyai tingkat akurasi yang signifikan ● Efisiensi dan biaya yang efektif, karena hanya menggunakan data historis (Hutasuhut et al., 2014)	● Tidak bisa menangkap hubungan non – linear ● Membutuhkan data yang stasioner ● Kurang sesuai untuk data musiman kompleks	● Digunakan dalam memprediksi deret waktu seperti curah hujan bulanan atau musiman berdasarkan data historis
ANN (*Artificial Neural Network)*	● Dapat menangani data kompleks dan non – linear	● Memerlukan data pelatihan yang besar	● Pengenalan pola dan pengambilan keputusan

Metode	Keunggulan	Kekurangan	Penggunaan Umum
	● Fleksibel jika digunakan dalam prediksi jangka panjang dan pendek ● Dapat mempelajari pola dari data (Watta & Hassoun, 1995)	● Rentan terhadap overfitting jika tidak diatur dengan benar ● Memerlukan waktu serta sumber daya komputasi yang besar (Haghshenas et al., 2020)	● Prediksi curah hujan yang menggunakan banyak variabel lingkungan seperti suhu, kelembaban, serta tekanan udara
LSTM (*Long Short-Term Memory)*	● Mampu mengatasi kelemahan RNN dalam menangani data urutan yang panjang ● Dapat menangkap informasi dari waktu sebelumnya dalam data deret waktu	● Lebih rumit dari jaringan syaraf biasa ● Waktu dalam proses pelatihan lama dan intensif ● Membutuhkan banyak sumber daya komputasi	● Prediksi deret waktu jangka panjang ● Prediksi curah hujan harian atau jam – jaman dengan menggunakan data sekuensial seperti kondisi atmosfer
Decision Tree	● Mudah dipahami serta diinterpretasikan ● Dapat menangani data numerik serta kategorikal ● Tidak memerlukan asumsi tentang distribusi data (Rokach & Maimon, 2006)	● Rentan terhadap overfitting tanpa pemangkasan yang tepat ● Kurang efektif dalam menangani interaksi yang komplek antar variabel (Rokach & Maimon, 2006)	● Regresi dan klasifikasi sederhana
Random Forest	● Mampu mengurangi overfitting dibandingkan dengan *decision tree* ● Mampu menangani data berdimensi tinggi dengan variabel yang banyak (Khan et al., 2021)	● Kompleksitas model membuat interpretasi hasil menjadi lebih sulit ● Membutuhkan waktu komputasi yang lama jika menggunakan dataset besar	● Digunakan untuk memprediksi data deret waktu atau klasifikasi dengan banyak variabel

Dalam penelitian ini, metode ARIMA (*Autoregressive Integrated Moving Average*) digunakan sebagai model utama dalam memprediksi curah hujan. Penggunaan metode ARIMA (*Autoregressive Integrated Moving Average*) dalam memprediksi curah hujan didasarkan pada sejumlah alasan ilmiah yang kuat. Pertama, ARIMA (*Autoregressive Integrated Moving Average*) efektif dalam menangani data deret waktu yang tidak stasioner, seperti curah hujan yang sering menunjukan fluktuasi musiman dan trend. Melalui proses differencing, ARIMA (*Autoregressive Integrated Moving Average*) Mampu mengubah data yang tidak stasioner menjadi stasioner, sehingga mempermudah dalam proses analisis lebih lanjut. Selain itu, metode ini menggabungkan elemen autoregressive (AR), moving average (MA), dan differencing, yang memungkinkan ARIMA (*Autoregressive Integrated Moving Average*) menangkap pola data historis untuk menghasilkan prediksi yang akurat. Efisiensi metode ARIMA (*Autoregressive Integrated Moving Average*) juga menjadi keunggulan yang signifikan, karena metode ini hanya memerlukan data historis tanpa memerlukan variabel – variabel eksternal tambahan, menjadikannya pilihan yang hemat waktu serta biaya dalam memprediksi curah hujan. Sebagai perbandingan metode - metode seperti ANN, LSTM, dan Random Forest yang disebutkan di sini sebagai metode alternatif dalam prediksi deret waktu, cenderung lebih kompleks dan membutuhkan data pelatihan yang besar, sehingga dalam penelitian ini, fokus utama tetap pada analisis dan implementasi metode ARIMA.

Fleksibilitas ARIMA (*Autoregressive Integrated Moving Average*)) dalam mengatasi berbagai jenis deret waktu, baik dengan trend musiman maupun fluktuasi jangka pendek, membuatnya sangat relevan dalam memprediksi curah hujan. Meskipun ARIMA (*Autoregressive Integrated Moving Average*) lebih optimal dalam menangani data linear, metode ini memberikan pendekatan yang valid serta sistematis untuk memprediksi trend jangka pendek maupun panjang. ARIMA (*Autoregressive Integrated Moving Average*) juga lebih tahan terhadap overfitting dibandingkan dengan Decision Tree, yang rentan terhadap masalah ini jika struktur pohon terlalu dalam. Keunggulannya dalam menyesuaikan parameter model dengan karakter data yang dianalisis memberikan kontribusi signifikan terhadap peningkatan akurasi prediksi dalam kerangka penelitian ini. Dengan demikian ARIMA (*Autoregressive Integrated Moving Average*) menjadi salah satu metode yang diandalkan dalam pemodelan deret waktu, terutama ketika pola curah hujan memiliki trend yang jelas dan pola musiman yang konsisten.

BAB III
METODOLOGI PENELITIAN

3.1 Tahapan ARIMA

Pada Gambar 3.1 menjelaskan tentang diagram alir tahapan ARIMA.



Gambar 3.1 Tahapan ARIMA

Pada Gambar 3.1 menjelaskan tentang proses ARIMA. Setiap tahapan dalam proses ini memiliki peran penting dalam memastikan bahwa model yang dihasilkan akurat. Pengambilan data dilakukan dengan teliti untuk memastikan data yang relevan dan berkualitas, diikuti dengan processing data untuk menghilangkan outlier dan mengisi nilai yang hilang atau kosong. Analisis awal data yang berisikan proses input data, lalu memastikan apakah data sudah stasioner atau belum. Proses selanjutnya adalah plot ACF dan PACF dengan cara melihat autokorelasi dan lag pada ACF dan PACF. Tahapan selanjutnya melakukan differencing yang tujuannya untuk menampilkan data yang sudah stasioner dengan menghilangkan trend.

Validasi model yang berisikan model-model ARIMA. Proses ini dilakukan untuk memastikan bahwa model tersebut mampu digunakan untuk memprediksi curah hujan dengan akurat menggunakan data yang telah disiapkan sebelumnya. Tahapan terakhir evaluasi berguna untuk menilai seberapa baik model dapat memprediksi curah hujan di masa depan.

3.2 Pengambilan Data

Pada tahap yang pertama dalam penelitian ini adalah pengumpulan data wilayah Sleman untuk (DIY), Semarang untuk (Jawa Tengah), dan Surabaya untuk (Jawa Timur) dari website data online BMKG (https://dataonline.bmkg.go.id/). Tahap ini melibatkan proses mengumpulkan informasi curah hujan yang mencakup jumlah hujan yang tercatat, jadwal pencatatan, dan mungkin faktor – faktor lain yang relevan. Data yang akurat dan komprehensif sangat penting karena menjadi dasar dari semua analisis yang akan dilakukan. Dalam pengumpulan data ini harus dilakukan dengan teliti agar memastikan keandalan hasil analisis dan prediksi yang akan dibuat nantinya. Data yang digunakan dalam penelitian ini diambil dari pencatatan curah hujan harian pada stasiun cuaca yang berada di wilayah Kabupaten Sleman, Semarang, dan Surabaya selama rentang waktu bulan Januari tahun 2020 hingga bulan Desember Tahun 2022. Data yang diperoleh sebanyak 3.288 data, terdapat beberapa nilai yang tidak lengkap atau tidak terukur. Nilai yang ditandai dengan kode 8888 menunjukkan nilai yang tidak terukur atau tidak tersedia. Selain itu, terdapat juga nilai null yang menandakan kekosongan data, serta kode 9999 yang menunjukkan ketiadaan data pada hari tersebut.

a. Data Curah Hujan Wilayah Kabupaten Sleman

Data curah hujan bulanan untuk Kabupaten Sleman bersumber dari Stasiun Klimatologi DI Yogyakarta. Periode pengumpulan data meliputi tiga tahun, dari tahun 2020 hingga 2022, dengan cakupan wilayah seluruh Kabupaten Sleman. Data tersebut disajikan dalam format bulanan.

Tabel 3.1 Contoh data curah hujan wilayah Kabupaten Sleman

Tanggal	Rainfall Rate (RR)
01-01-2020	29,5
02-01-2020	30,9
03-01-2020	13,3
04-01-2020	46,6
05-01-2020	3
06-01-2020	
07-01-2020	29,6
08-01-2020	0
09-01-2020	2,6

b. Data Curah Hujan Wilayah Semarang

Stasiun Klimatologi Jawa Tengah menjadi sumber data curah hujan bulanan Kota Semarang. Data ini dikumpulkan dalam format bulanan untuk periode tiga tahun, dari tahun 2020 hingga 2022, dan difokuskan pada wilayah Kota Semarang.

Tabel 3.2 Contoh data curah hujan wilayah Kota Semarang

Tanggal	Rainfall Rate (RR)
01-01-2020	9,6
02-01-2020	16,7
03-01-2020	2
04-01-2020	36,6
05-01-2020	3,7
06-01-2020	20,3
07-01-2020	28,8
08-01-2020	28
09-01-2020	13,5

c. Data Curah Hujan Wilayah Surabaya

Stasiun Meteorologi Perak I menyediakan data curah hujan bulanan khusus untuk wilayah Surabaya. Data ini dikumpulkan dalam format bulanan untuk periode tiga tahun, dari tahun 2020 hingga 2022.

Tabel 3.3 Contoh data curah hujan wilayah Surabaya

Tanggal	Rainfall Rate (RR)
01-01-2020	8888
02-01-2020	1,2
03-01-2020	0
04-01-2020	5,4
05-01-2020	1,4
06-01-2020	16,2
07-01-2020	58,5
08-01-2020	1,5
09-01-2020	10

3.3 Processing Data

Tahap selanjutnya dalam melakukan prediksi curah hujan adalah processing data. Tahap ini merupakan tahap penting yang akan melibatkan proses utama untuk mempersiapkan data sebelum dianalisis lebih lanjut. Proses ini dimulai dari pembersihan data, dimana data yang telah dikumpulkan diperiksa untuk nilai yang hilang, duplikat atau outlier yang tidak biasa. Tahap ini krusial untuk memastikan kebersihan dan keakuratan data yang diperoleh sebelum nantinya data digunakan dalam analisis lebih lanjut.

a. Pemeriksaan Data Awal

Pada penelitian ini dilakukan identifikasi terhadap data yang telah diperoleh, apakah ditemukan nilai null, nilai tidak terukur (8888), dan tidak ada data (9999). Hal ini penting untuk memastikan bahwa tidak ada data yang hilang yang dapat mempengaruhi hasil analisis. Berikut ini contoh sederhana dari proses pemeriksaan awal data.

Tabel 3.4 Contoh pemeriksaan data awal

Tanggal	Rainfall Rate (RR)
01-01-2020	29,5
02-01-2020	9999
03-01-2020	8888
04-01-2020	46,6
05-01-2020	
06-01-2020	16,2
07-01-2020	58,5

Tanggal	Rainfall Rate (RR)
08-01-2020	1,5
09-01-2020	
10-01-2020	37

Pada tahap ini, peneliti mengidentifikasi data yang bermasalah, yaitu:

Nilai “9999” pada tanggal 02-01-2020 (tidak ada data).

Nilai “8888” pada tanggal 03-01-2020 (nilai tidak terukur).

Nilai “null” pada tanggal 05-01-2020 dan 09-01-2020 (nilai kosong).

b. Penggantian Nilai Kosong dengan Median

Setelah proses pemeriksaan awal data, dilakukan penggantian nilai null, nilai tidak terukur (8888), dan tidak ada data (9999) dengan nilai median. Nilai median dipilih karena lebih tahan terhadap outlier dibandingkan nilai rata – rata, sehingga menghasilkan data yang lebih akurat dan terhindar dari distorsi. Berikut ini Contoh Sederhana dari proses penggantian nilai kosong dengan median.

Tabel 3.5 Contoh penggantian nilai kosong dengan median

Tanggal	Rainfall Rate (RR)
01-01-2020	29,5
02-01-2020	9999
03-01-2020	8888
04-01-2020	46,6
05-01-2020	
06-01-2020	16,2
07-01-2020	58,5
08-01-2020	1,5
09-01-2020	
10-01-2020	37

Pada tahap ini, Peneliti melakukan perhitungan nilai median dari data yang didapat tanpa memasukan nilai (“9999”, “8888”, “null”):

Data valid: 29.5, 46.6, 16.2, 58.5, 1.5, 37.

Urutkan data: 1.5, 16.2, 29.5, 37, 46.6, 58.5.

Median (nilai tengah):

Me = Xn/2 + X(X/2+1) / 2

Me = 29.5 + 37 / 2

Me = 66.5 / 2

Me = 33.25

Setelah mendapatkan nilai median nya, peneliti mengganti nilai yang bermasalah pada data yang sudah di dapat.

Tabel 3.6 Contoh penggantian nilai setelah mendapatkan median

Tanggal	**Rainfall Rate (RR)**
01-01-2020	29,5
02-01-2020	33.25
03-01-2020	33.25
04-01-2020	46,6
05-01-2020	33.25
06-01-2020	16,2
07-01-2020	58,5
08-01-2020	1,5
09-01-2020	33.25
10-01-2020	37

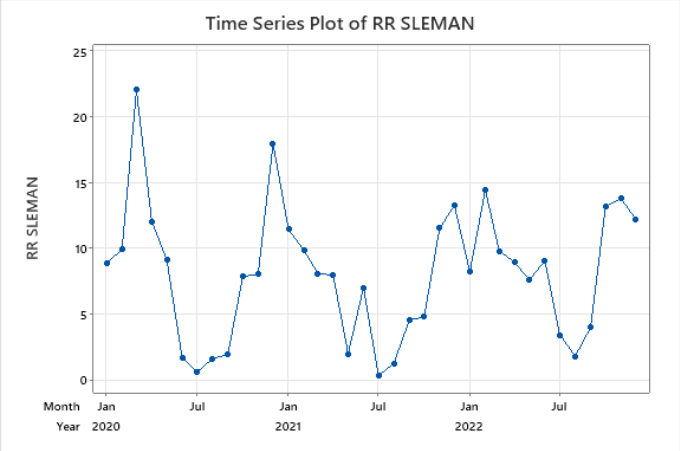
3.4 Analisis Awal Data

Tahap selanjutnya adalah analisis awal data. Di tahap ini, visualisasi data curah hujan dilakukan untuk mengidentifikasi pola atau tren yang ada dalam data. Dengan membuat grafik, diagram, atau statistik deskriptif, peneliti dapat memahami lebih dalam distribusi curah hujan dari waktu ke waktu atau menemukan pola lain yang mungkin tersembunyi dalam data. Analisis ini memberikan pemahaman awal yang penting tentang karakteristik utama data sebelum memasuki fase pengembangan model prediksi.

3.4.1Plot Time Series

Pada tahap pertama dalam analisis awal data adalah melakukan plot time series. Plot time series adalah cara yang efektif untuk memvisualisasikan data curah hujan, dengan plot time series dapat mengidentifikasi pola atau tren dari data curah hujan yang telah didapatkan.

a. Pada Gambar 3.2 menjelaskan tentang Plot Time Series Wilayah Kabupaten Sleman

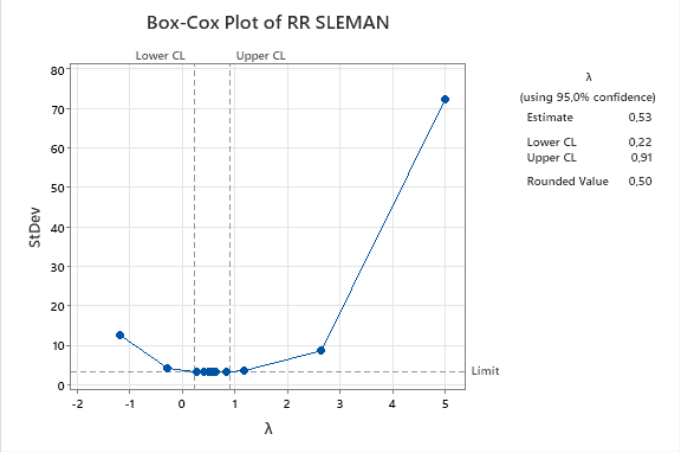


Gambar 3.2 Contoh Plot Time Series Kabupaten Sleman

3.4.2Uji Stasioneritas

Setelah melakukan tahap plot time series, tahap selanjutnya melakukan uji stasioner dari data curah hujan yang didapatkan apakah data curah hujan tersebut sudah stasioner atau belum stasioner.

a. Pada Gambar 3.3 menjelaskan tentang Uji stasioneritas wilayah Kabupaten Sleman

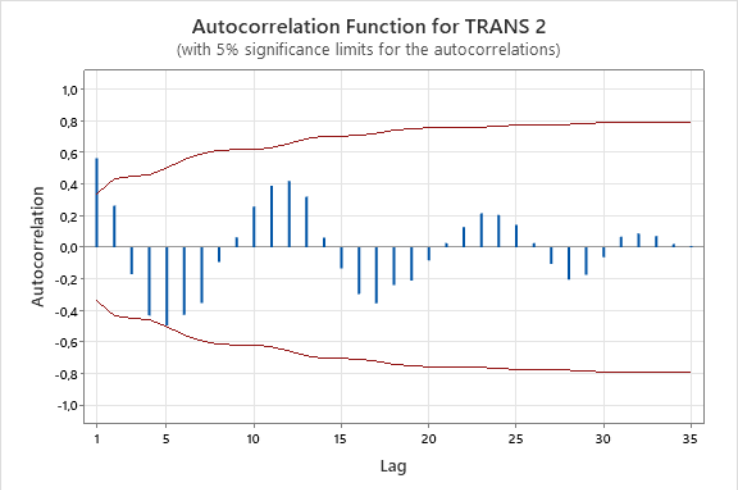


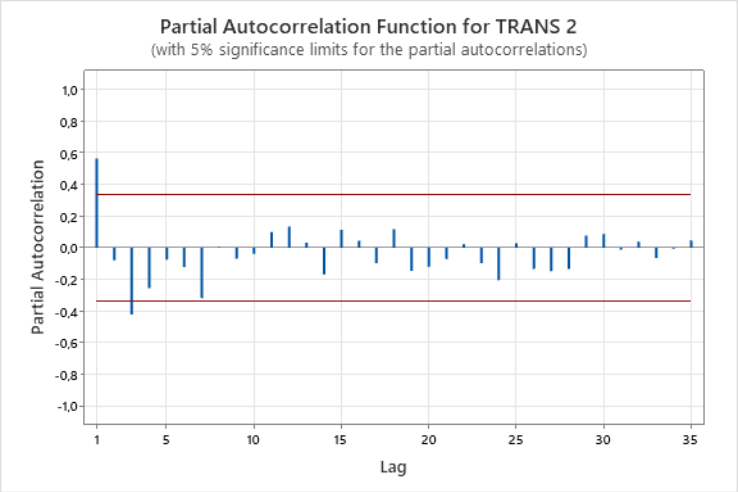
Gambar 3.3 Contoh Uji Stasioneritas Kabupaten Sleman

3.4.3Plot ACF dan PACF

ACF (Autokorelasi) adalah proses dimana mengukur korelasi antara nilai data pada waktu tertentu dengan nilai data pada waktu sebelumnya. ACF dihitung untuk berbagai lag, yaitu selisih antara dua pengamatan. Sebagai contoh nilai ACF yang tinggi di lag 1 (0,123) dan lag 2 (0,491) menunjukan bahwa waktu saat ini memiliki korelasi yang kuat dengan waktu sebelumnya. Nilai ACF yang lebih rendah di lag selanjutnya menunjukan bahwa korelasi semakin lemah seiring dengan bertambahnya lag. Sedangkan PACF (Autokorelasi Parsial) adalah proses pengukuran korelasi secara parsial antara nilai data pada waktu tertentu dengan nilai data pada waktu sebelumnya. Sebagai contoh nilai PACF tertinggi ada pada lag 1 (0,123) dan lag 2 (0,368) menunjukan bahwa waktu saat ini memiliki korelasi parsial yang kuat dengan waktu sebelumnya. Nilai PACF yang lebih rendah di lag selanjutnya menunjukan bahwa korelasi parsial semakin lemah seiring dengan bertambahnya lag. PACF memfokuskan pada korelasi tanpa mempertimbangkan pengaruh nilai data pada lag yang lebih kecil.

a. Pada Gambar 3.4 menjelaskan tentang ACF dan PACF wilayah Kabupaten Sleman



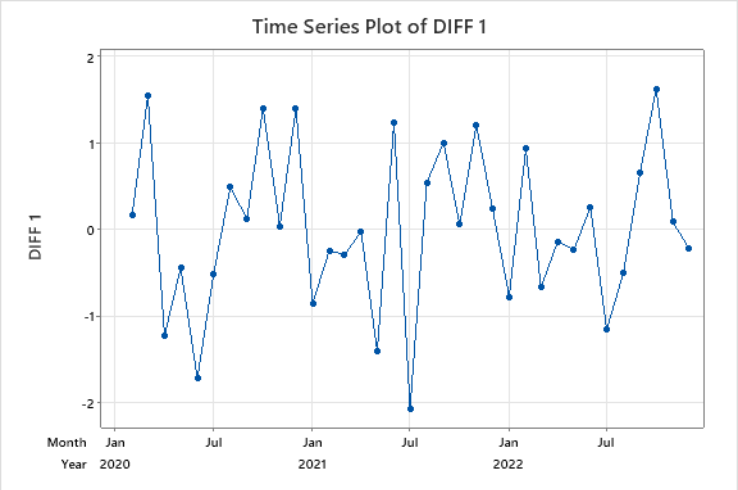


Gambar 3.4 Contoh ACF dan PACF Kabupaten Sleman

3.4.4Differencing

Tahap selanjutnya adalah differencing, tahap ini berguna untuk menampilkan data yang sudah stasioner dengan menghilangkan trend atau pola data curah hujan.

a. Pada Gambar 3.5 menjelaskan tentang Differencing wilayah Kabupaten Sleman



Gambar 3.5 Contoh Differencing Kabupaten Sleman

3.5 Validasi Model

Setelah analisis awal data, langkah berikutnya adalah validasi model. Data yang telah diproses digunakan untuk membuat model prediksi curah hujan. Validasi model memastikan bahwa model tersebut mampu memprediksi curah hujan dengan akurat menggunakan data yang telah disiapkan sebelumnya. Tahap ini sangat penting untuk memastikan bahwa model yang akan digunakan untuk prediksi telah teruji dan terverifikasi keandalannya.

3.6 Evaluasi Model

Tahap terakhir adalah evaluasi model. Setelah validasi, model dievaluasi untuk mengukur kinerjanya menggunakan metrik seperti MSE. *Mean Squared Error* (MSE) adalah suatu matrik evaluasi matrix yang umum digunakan untuk mengukur kinerja model, termasuk ARIMA dalam memprediksi curah hujan. Evaluasi ini membantu menilai seberapa baik model dapat memprediksi curah hujan dengan membandingkan hasil prediksi dengan nilai sebenarnya. Tahap evaluasi ini krusial karena memastikan bahwa model yang telah dibangun dari data yang telah diproses dapat memberikan hasil yang handal dan berguna bagi analisis dan prediksi kedepannya. Selain MSE terdapat juga MAPE. *Mean Absolute Percentage Error* (MAPE) adalah suatu matrik evaluasi yang berguna untuk mengevaluasi kinerja model. Metrik ini mudah digunakan karena memberikan hasil dalam bentuk persentase yang membuatkan mudah untuk diinterpretasikan.

BAB IV
HASIL DAN PEMBAHASAN

4.1 Detail Spesifikasi Perangkat Keras dan Lunak yang Digunakan

Perangkat keras dan perangkat lunak yang diperlukan untuk melakukan penelitian ini akan diuraikan di bawah ini.

4.1.1 Perangkat Keras (*Hardware*)

1. Laptop Asus ROG Strix G531GD
2. Processor Intel® Core™ i5-9300H CPU @ 2.40GHz (8 CPUs), ~2.4GHz
3. NVDIA GeForce GTX 1050
4. Memory 16384MB RAM

4.1.2 Perangkat Lunak (*Software*)

1. System Operasi Windows 11 Home Single Language 64-bit
2. Microsoft Excel 2019
3. Minitab® 21.4.2 (64-bit)

Penggunaan Laptop Asus ROG Strix G531GD memastikan bahwa proses analisis data dapat dilakukan dengan lancar dan tanpa hambatan performa. Dalam tahap awal, data curah hujan yang dikumpulkan diolah menggunakan perangkat lunak Microsoft Excel untuk mengatur dan menyiapkan data sebelum masuk ke tahap analisis lebih lanjut. Selain itu, untuk tahap analisis statistik dan pemodelan ARIMA, perangkat lunak Minitab digunakan sebagai alat utama. Minitab memberikan berbagai fitur analisis statistik yang komprehensif dan memungkinkan penyusunan model ARIMA dengan akurat, yang merupakan bagian inti dari penelitian ini.

4.2 Analisis Awal Data

Data curah hujan yang telah diperiksa dan sudah bersih selanjutnya mencari karakteristik curah hujan pada tiap stasiun cuaca yang berada di wilayah Kabupaten Sleman, Semarang, dan Surabaya.

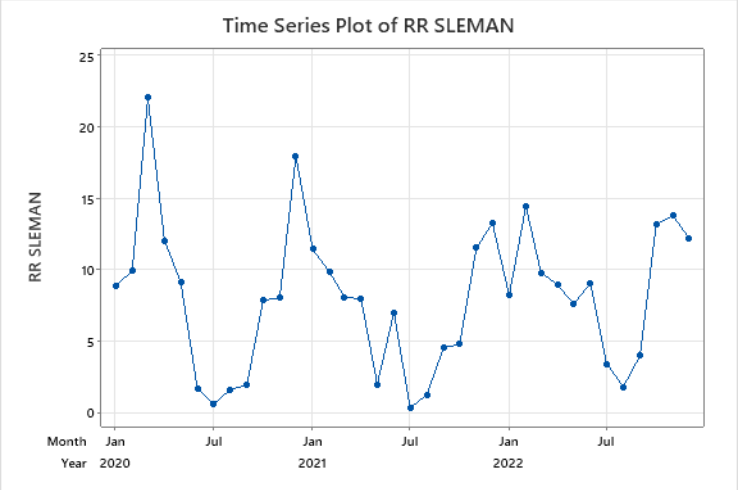
Tabel 4.1 Ringkasan Data Curah Hujan

Analisis Ringkasan	Stasiun Klimatologi DI Yogyakarta	Stasiun Klimatologi Jawa Tengah	Stasiun Meteorologi Perak I
Nilai Minimum (mm)	0,00	0,00	0,00
Nilai Maximum (mm)	230,50	173,50	133,90
Mean (mm)	8,04	6,97	6,46
Median (mm)	1,35	0,30	0,30

4.2.1Plot Time Series

Setelah tahap mencari karakteristik data curah hujan di ketiga wilayah, tahap selanjutnya adalah memvisualisasikan data curah hujan tiap wilayah dengan menggunakan plot time series sebagai berikut.

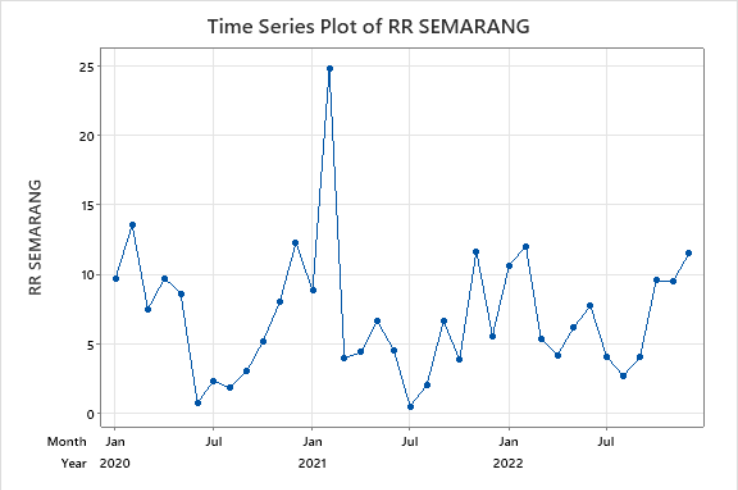
a. Wilayah Kabupaten Sleman



Gambar 4.1 Plot Time Series Kabupaten Sleman

Menurut Gambar 4.1, nilai maximum yang terdapat pada Stasiun Klimatologi DI Yogyakarta pada tahun 2020 sebesar 230,5mm, tahun 2021 sebesar 84,8mm, dan pada tahun 2022 sebesar 97,6mm.

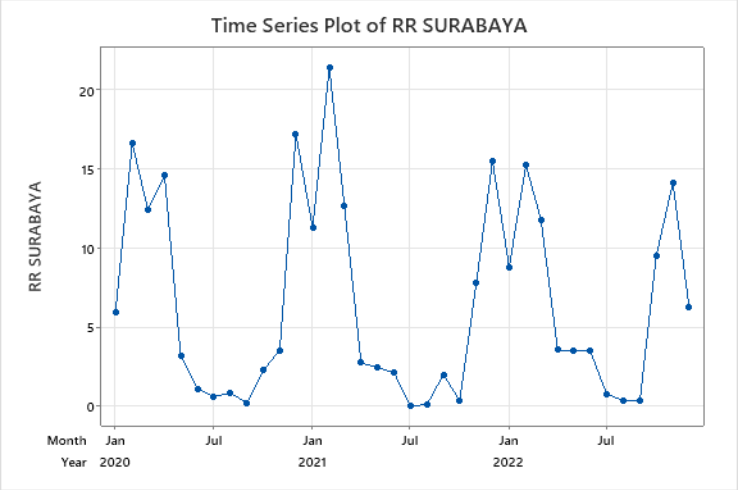
b. Wilayah Semarang



Gambar 4.2 Plot Time Series Semarang

Menurut Gambar 4.2, nilai maximum yang terdapat pada Stasiun Klimatologi Jawa Tengah pada tahun 2020 sebesar 114mm, tahun 2021 sebesar 173,5mm, dan pada tahun 2022 sebesar 95,3mm.

c. Wilayah Surabaya

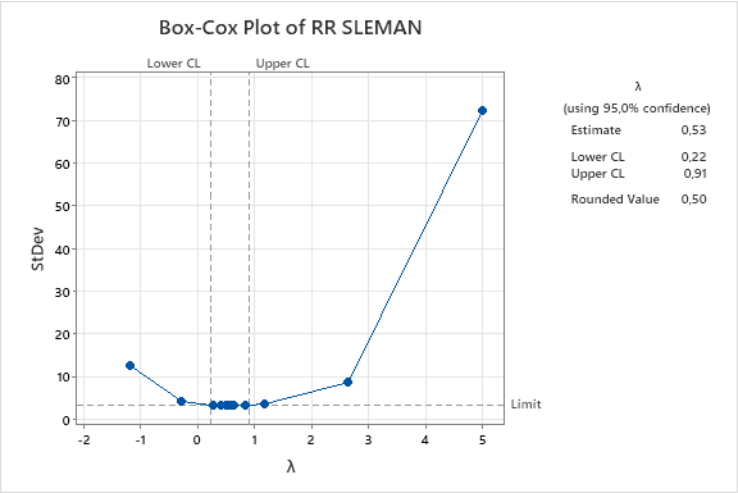


Gambar 4.3 Plot Time Series Surabaya

Menurut Gambar 4.3, nilai maximum yang terdapat pada Stasiun Meteorologi Perak I pada tahun 2020 sebesar 133,9mm, tahun 2021 sebesar 102,4mm, dan pada tahun 2022 sebesar 87,7mm.

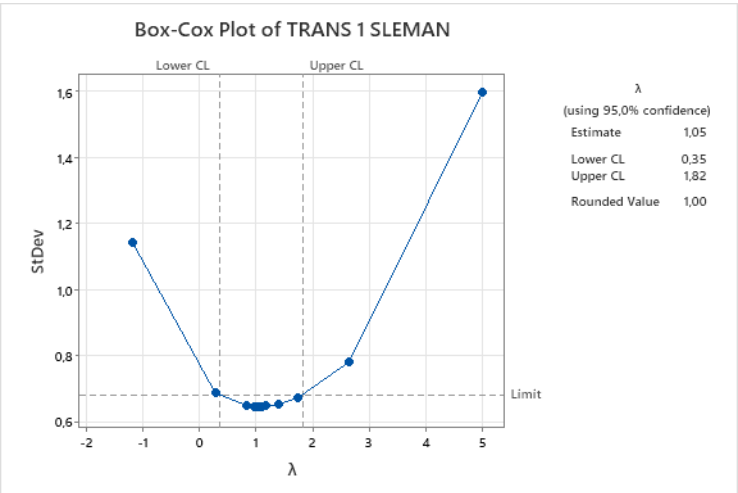
4.2.2Uji Stasioneritas

a. Uji stasioneritas wilayah Kabupaten Sleman



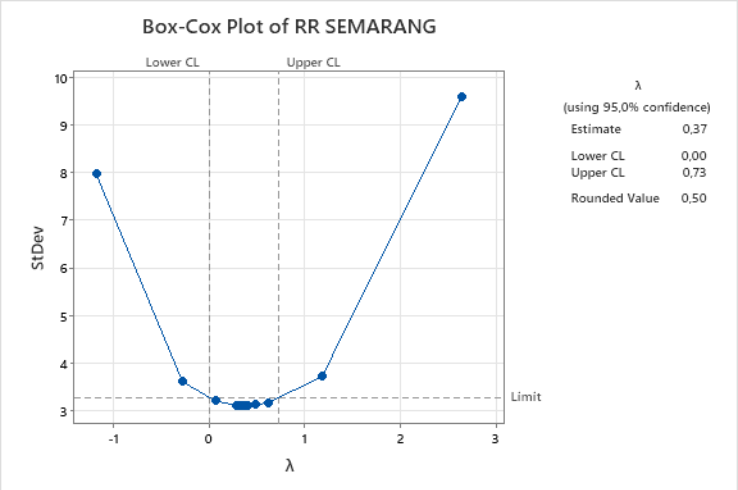
Gambar 4.4 Nilai Uji Stasioneritas Kabupaten Sleman Sebelum Transformasi

Menurut Gambar 4.4, nilai uji stasioneritas data curah hujan wilayah Kabupaten Sleman menunjukan standar deviasi sebesar λ 0,50. Dari hasil tersebut data curah hujan wilayah Kabupaten Sleman mempunyai variasi yang cukup rasional. Hal ini di sebabkan oleh data yang tersebar luas ataupun terdapat sebagian nilai ekstrem dalam data curah hujan wilayah Kabupaten Sleman, sehingga perlu dilakukan transformasi agar nilai λ menjadi satu, dengan mentransformasi nilai deviasi menjadi satu data akan lebih seragam dan lebih mudah dalam proses analisis. Berikut hasil transformasi Box-Cox Plot wilayah Kabupaten Sleman.



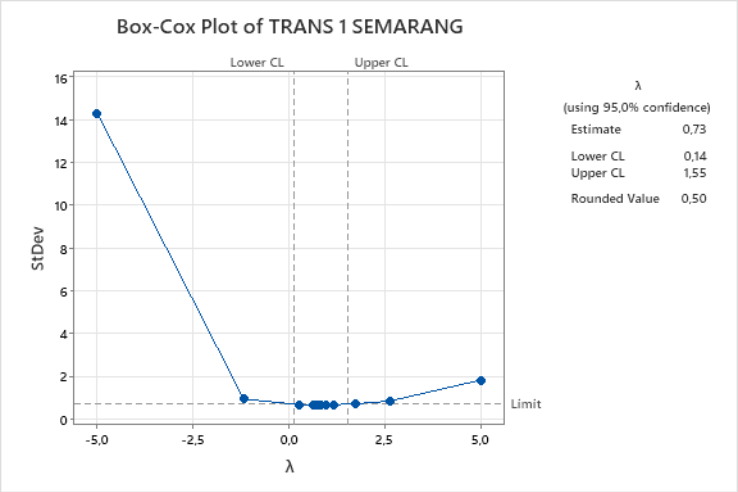
Gambar 4.5 Nilai Uji Stasioneritas Kabupaten Sleman Setelah Transformasi

b. Uji stasioneritas wilayah Semarang

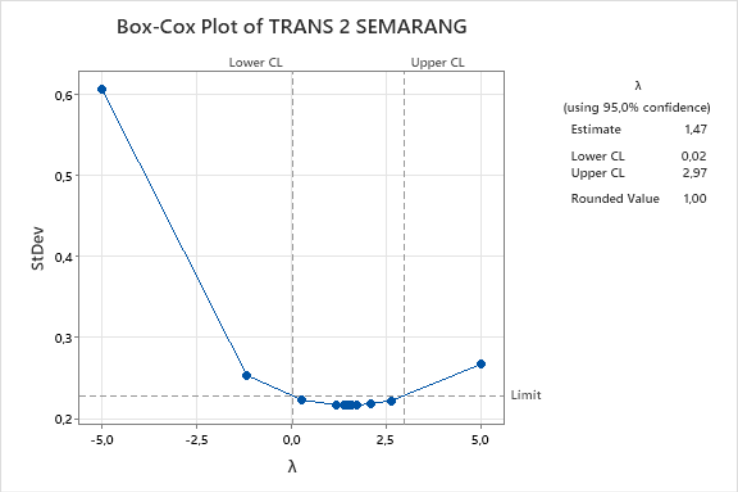


Gambar 4.6 Nilai Uji Stasioneritas Semarang Sebelum Transformasi

Menurut Gambar 4.6, nilai uji stasioneritas data curah hujan Semarang menunjukan nilai standar deviasi sebesar λ 0,50. Dibandingkan dengan data wilayah Kabupaten Sleman, data wilayah Semarang lebih konsisten, yang menunjukkan sebaran data lebih ketat di sekitar rata-rata, namun tetap ada variasi yang membutuhkan transformasi untuk mencapai standar deviasi yang diinginkan. Nilai standar deviasi sebesar 0.50 setelah satu kali transformasi menunjukkan bahwa data masih memiliki sedikit variasi yang tidak sepenuhnya hilang. Hal ini bisa disebabkan oleh adanya sedikit outlier atau ketidakhomogenan dalam data yang tetap ada setelah transformasi pertama. Oleh karena itu, diperlukan transformasi tambahan untuk lebih meratakan data dan mencapai standar deviasi yang diharapkan, yaitu satu. Berikut hasil transformasi Box-Cox Plot wilayah Semarang.

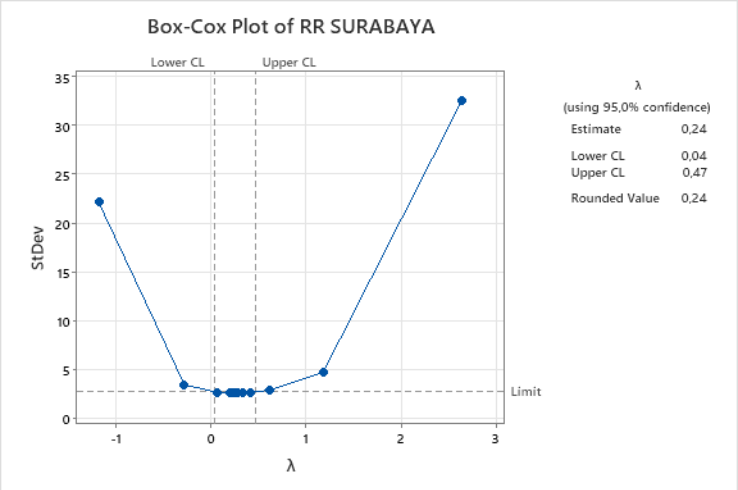


Gambar 4.7 Nilai Uji Stasioneritas Semarang Setelah Transformasi Satu Kali



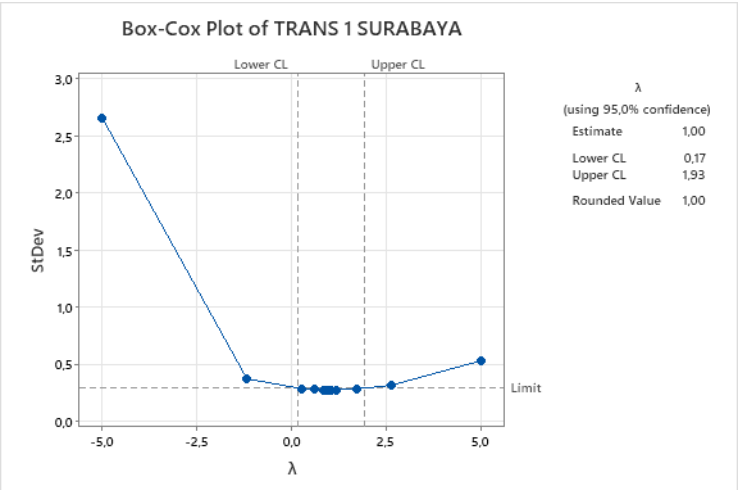
Gambar 4.8 Nilai Uji Stasioneritas Semarang Setelah Transformasi Dua Kali

c. Uji stasioneritas wilayah Surabaya



Gambar 4.9 Nilai Uji Stasioneritas Surabaya Sebelum Transformasi

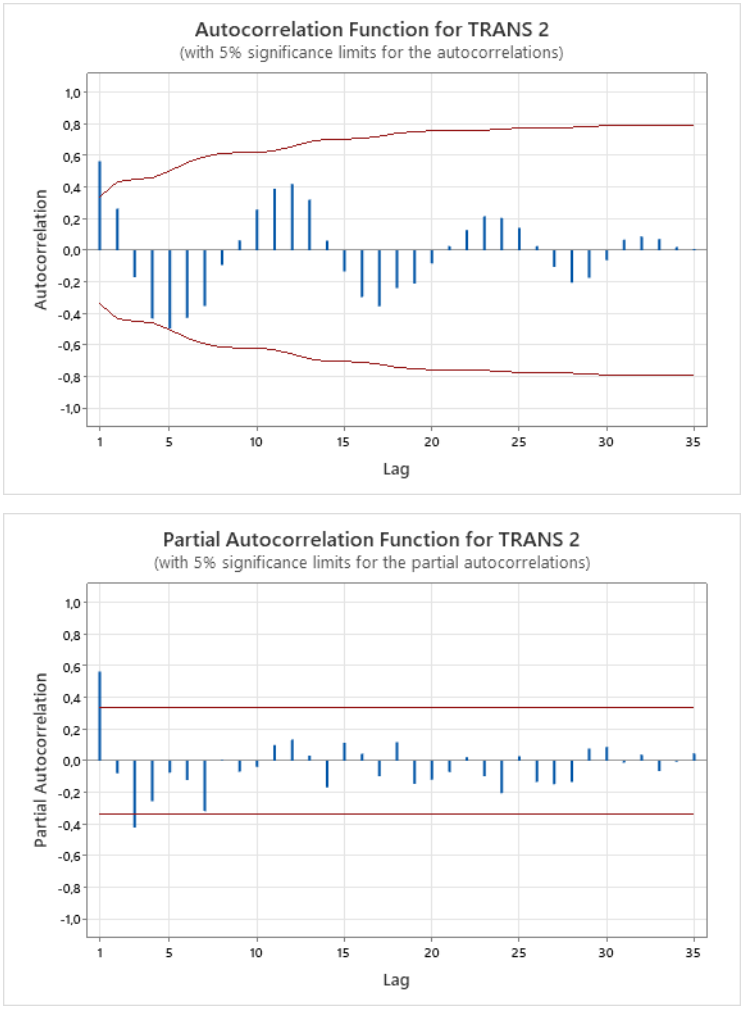
Menurut Gambar 4.9, nilai uji stasioneritas data curah hujan wilayah Surabaya menunjukan nilai standar deviasi sebesar λ 0,24. Data curah hujan wilayah Surabaya memiliki variasi yang paling kecil dibandingkan dengan wilayah Kabupaten Sleman dan Semarang. Data wilayah Surabaya lebih homogen dan tersebar lebih dekat ke rata-ratanya. Ini menunjukkan bahwa data curah hujan wilayah Surabaya lebih konsisten dan memiliki lebih sedikit variasi antar nilai datanya, sehingga lebih mudah untuk ditransformasi. Dari hasil tersebut data curah hujan wilayah Surabaya perlu dilakukan transformasi agar nilai λ menjadi satu. Dengan mentransformasi nilai deviasi menjadi satu data akan lebih seragam dan lebih mudah dalam proses analisis. Berikut hasil transformasi Box-Cox Plot wilayah Surabaya.



Gambar 4.10 Nilai Uji Stasioneritas Surabaya Setelah Transformasi

4.2.3Plot ACF dan PACF

a. ACF dan PACF wilayah Kabupaten Sleman

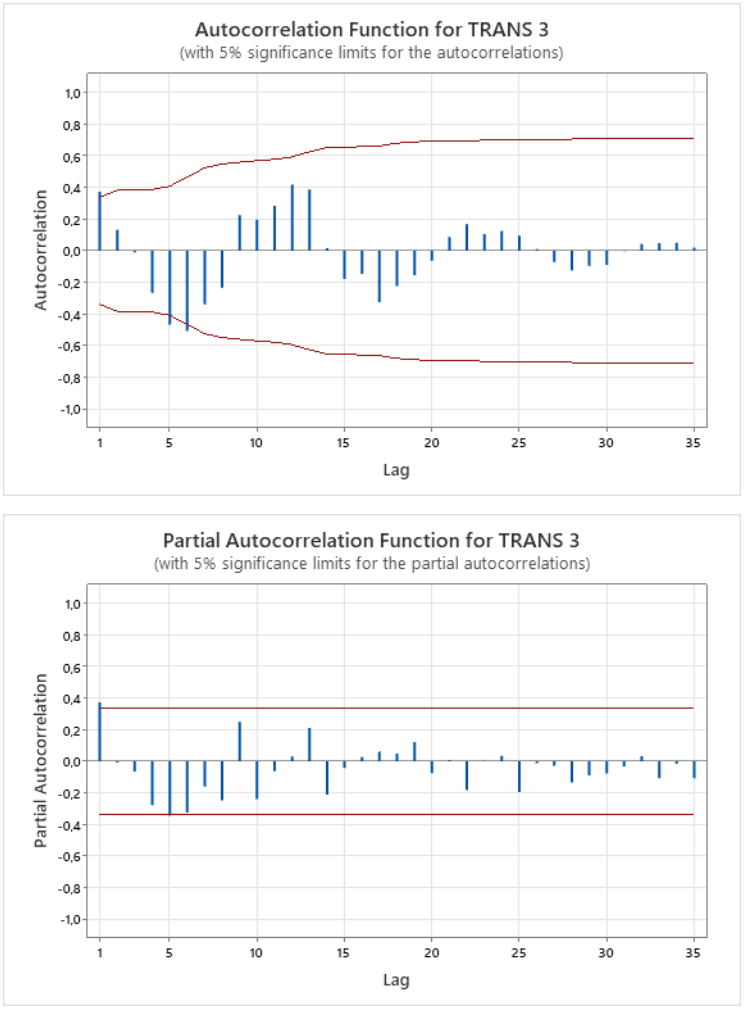


Gambar 4.11 ACF dan PACF Kabupaten Sleman

Berdasarkan Gambar 4.11, fungsi autokorelasi (ACF) dan fungsi autokorelasi parsial (PACF) untuk wilayah Kabupaten Sleman menunjukkan pola tertentu karena beberapa faktor utama. Pada grafik ACF, terlihat bahwa autokorelasi signifikan pada beberapa lag awal, terutama hingga lag ke-10, yang menunjukkan adanya pola musiman atau tren dalam data. Penurunan autokorelasi pada ACF terjadi secara lambat, menunjukkan adanya komponen non-stasioner dalam data. Pola musiman, seperti variasi bulanan atau tahunan, juga tercermin dalam puncak-puncak autokorelasi pada lag tertentu. Kejadian khusus atau perubahan besar, seperti perubahan kebijakan, dapat menyebabkan lonjakan atau penurunan tajam dalam data, menciptakan autokorelasi signifikan. Pada grafik PACF, autokorelasi parsial signifikan hanya pada beberapa lag awal, dan penurunan cepat pada PACF ini menunjukkan bahwa setelah mengontrol efek lag sebelumnya, tidak banyak lag yang memiliki korelasi signifikan. Secara

keseluruhan, nilai-nilai data masa lalu masih mempengaruhi nilai-nilai data saat ini, meskipun pengaruhnya berkurang seiring waktu.

b. ACF dan PACF wilayah Semarang

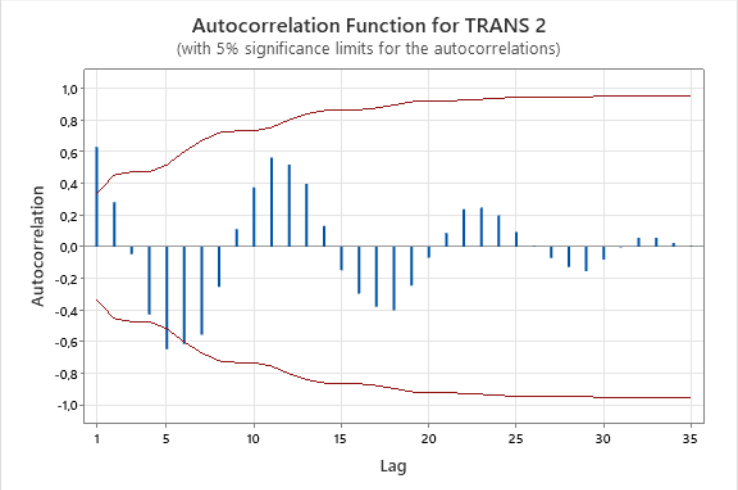


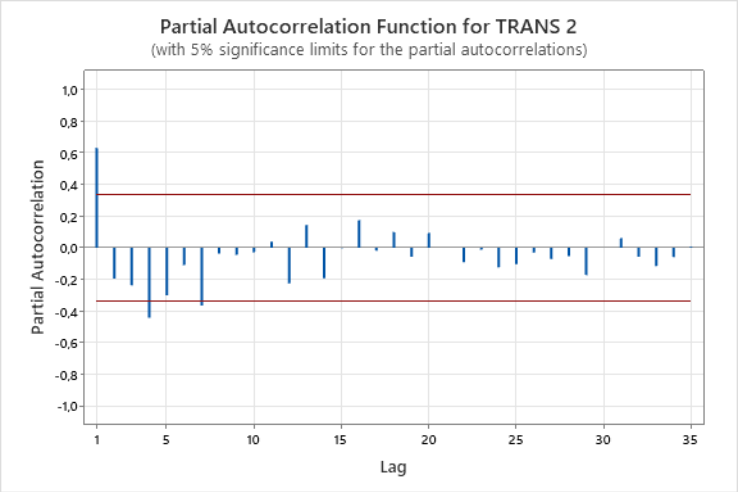
Gambar 4.12 ACF dan PACF Semarang

Berdasarkan Gambar 4.12 menunjukkan fungsi autokorelasi (ACF) dan fungsi autokorelasi parsial (PACF) untuk wilayah Semarang. Grafik ACF menunjukkan adanya autokorelasi signifikan pada lag-lag awal, khususnya hingga lag ke-10, dengan penurunan autokorelasi yang terjadi secara lambat, mirip dengan yang terlihat di wilayah Kabupaten Sleman. Ini menunjukkan adanya pola musiman atau tren dalam data. Pada grafik PACF, autokorelasi parsial juga terlihat signifikan pada beberapa lag awal, dengan penurunan cepat pada PACF, yang menunjukkan bahwa setelah mengontrol efek lag sebelumnya, tidak banyak lag yang memiliki korelasi signifikan.

Hasil ACF dan PACF dari data wilayah Semarang menunjukkan pola tertentu karena beberapa faktor utama. Tren dalam data, seperti peningkatan atau penurunan yang konsisten, menyebabkan nilai-nilai data masa lalu masih mempengaruhi nilai-nilai data saat ini, terlihat dari autokorelasi tinggi pada lag-lag awal. Pola musiman, seperti variasi bulanan atau tahunan, juga tercermin dalam puncak-puncak autokorelasi pada lag tertentu. Kejadian khusus atau perubahan besar, seperti perubahan kebijakan, dapat menyebabkan lonjakan atau penurunan tajam dalam data, menciptakan autokorelasi signifikan. Pada grafik PACF, autokorelasi parsial signifikan hanya pada beberapa lag awal, dan penurunan cepat pada PACF ini menunjukkan bahwa setelah mengontrol efek lag sebelumnya, tidak banyak lag yang memiliki korelasi signifikan. Secara keseluruhan, nilai-nilai data masa lalu masih mempengaruhi nilai-nilai data saat ini, meskipun pengaruhnya berkurang seiring waktu.

c. ACF dan PACF wilayah Surabaya





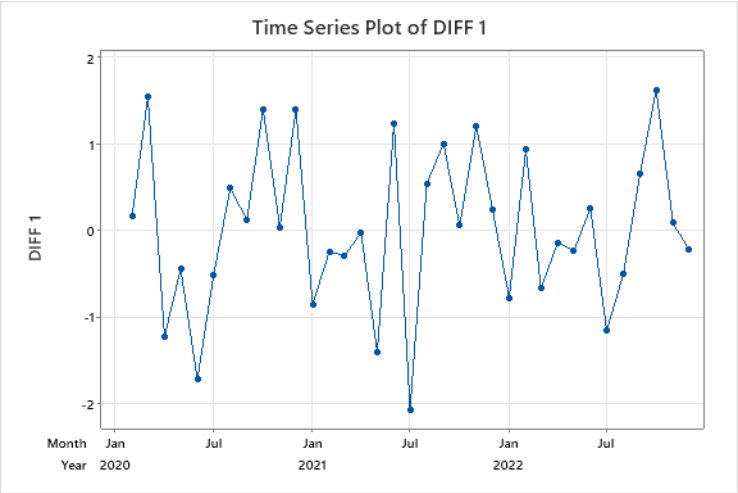
Gambar 4.13 ACF dan PACF Surabaya

Berdasarkan Gambar 4.13 menunjukkan fungsi autokorelasi (ACF) dan fungsi autokorelasi parsial (PACF) untuk wilayah Surabaya. Grafik ACF Pada grafik ACF, terlihat bahwa autokorelasi signifikan pada beberapa lag awal, terutama hingga lag ke-10, yang menunjukkan adanya pola musiman atau tren dalam data. Penurunan autokorelasi pada ACF terjadi secara lambat, menunjukkan adanya komponen non-stasioner dalam data. Pola musiman, seperti variasi bulanan atau tahunan, juga tercermin dalam puncak-puncak autokorelasi pada lag tertentu. Kejadian khusus atau perubahan besar, seperti perubahan kebijakan, dapat menyebabkan lonjakan atau penurunan tajam dalam data, menciptakan autokorelasi signifikan. Pada grafik PACF, autokorelasi parsial signifikan hanya pada beberapa lag awal, dan penurunan cepat pada PACF ini menunjukkan bahwa setelah mengontrol efek lag sebelumnya, tidak banyak lag yang memiliki korelasi signifikan. Secara keseluruhan, nilai-nilai data masa lalu masih mempengaruhi nilai-nilai data saat ini, meskipun pengaruhnya berkurang seiring waktu.

4.2.4Differencing

Setelah mendapatkan hasil nilai ACF dan PACF tahapan selanjutnya adalah differencing yang berfungsi untuk menghilangkan pola atau trend yang terdapat pada data.

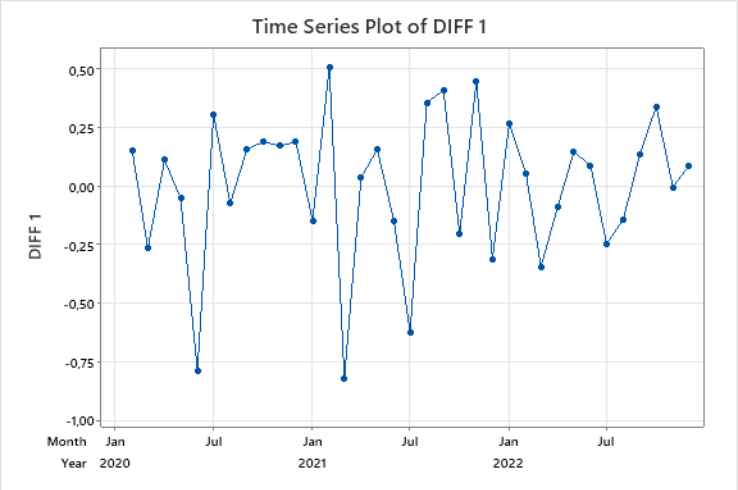
a. Differencing wilayah Kabupaten Sleman



Gambar 4.14 Differencing Kabupaten Sleman

Hasil penerapan differencing terhadap data curah hujan wilayah Kabupaten Sleman disajikan pada Gambar 4.14. Differencing adalah metode yang digunakan untuk menghilangkan trend atau pola musiman dalam data untuk mencapai stasioneritas. Grafik menampilkan rentang waktu Januari 2020 hingga Juli 2022 pada sumbu horizontal (X) dan nilai pembeda pertama pada sumbu vertikal (Y). Fluktuasi nilai differencing wilayah Kabupaten Sleman berada di sekitar garis nol menunjukkan keberhasilan pengurangan tren atau pola musiman pada data asli, sehingga menghasilkan kumpulan data yang lebih stabil.

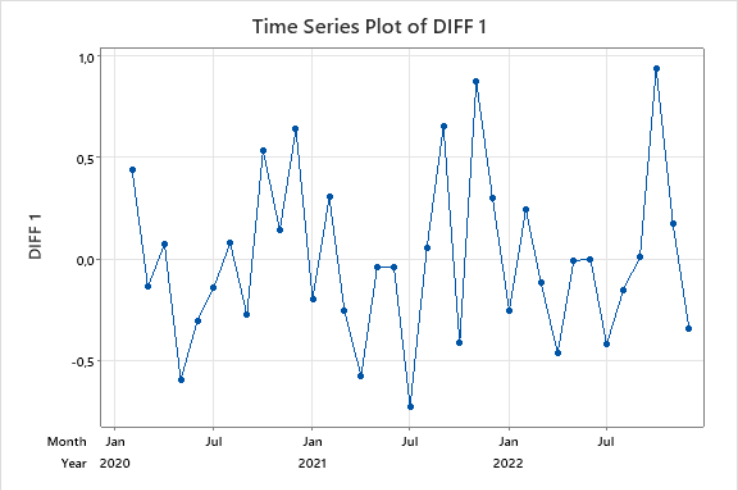
b. Differencing wilayah Semarang



Gambar 4.15 Differencing Semarang

Grafik yang disajikan pada Gambar 4.15 menggambarkan hasil penerapan teknik differencing terhadap data curah hujan wilayah Semarang. Differencing digunakan untuk menghilangkan tren atau pola dalam data, sehingga meningkatkan stasioneritasnya. Sumbu horizontal grafik mewakili rentang waktu dari Januari 2020 hingga Juli 2022, sedangkan sumbu vertikal menampilkan nilai yang dihasilkan dari pembedaan pertama. Osilasi nilai differencing wilayah Semarang yang konsisten di sekitar garis nol menunjukkan bahwa prosedur pembedaan telah secara efektif memitigasi tren atau pola musiman, yang mengarah pada peningkatan stasioneritas data.

c. Differencing wilayah Surabaya



Gambar 4.16 Differencing Surabaya

Data yang disajikan pada Gambar 4.16 mengilustrasikan hasil penerapan differencing terhadap data curah hujan wilayah Surabaya. Tujuan dari teknik differencing ini adalah untuk menghilangkan tren mendasar atau variasi musiman yang ada pada data asli, sehingga meningkatkan stasioneritasnya. Sumbu x mewakili jangka waktu dari Januari 2020 hingga Juli 2022, sedangkan sumbu y menampilkan nilai yang dihasilkan dari proses pembedaan pertama. Fluktuasi nilai differencing wilayah Surabaya yang berada di sekitar garis nol menunjukkan efektivitas pembedaan dalam memitigasi tren atau pola musiman yang ada, sehingga menghasilkan kumpulan data yang lebih stabil.

4.3 Validasi Model

a. Model Wilayah Kabupaten Sleman

Tabel 4.2 Model ARIMA Kabupaten Sleman

MODEL ARIMA	MSE
0, 1, 1	0,865332
1, 1, 1	0,878936
1, 1, 0	1,25034

Tabel 4.2 menyajikan hasil validasi model ARIMA untuk Kabupaten Sleman, yang menampilkan tiga konfigurasi model yang berbeda: ARIMA (0, 1, 1), ARIMA (1, 1, 1), dan ARIMA (1, 1, 0). Hasil validasi digambarkan melalui metrik *Mean Squared Error* (MSE) yang berfungsi sebagai indikator akurasi prediksi model. Di antara ketiga konfigurasi, model ARIMA (0, 1, 1) menunjukkan nilai MSE terendah yaitu 0,865332, yang menunjukkan kinerja

prediktif yang lebih unggul dibandingkan model lainnya. Sebaliknya, model ARIMA (1, 1, 0) menunjukkan nilai MSE tertinggi sebesar 1,25034, menunjukkan kemampuan prediksi yang kurang optimal. Pada akhirnya, model ARIMA (0, 1, 1) menjadi pilihan yang paling sesuai untuk prediksi data di Kabupaten Sleman karena menghasilkan kesalahan prediksi yang paling kecil. Hasil validasi menggarisbawahi kemanjuran model ARIMA dengan parameter p=0, d=1, dan q=1 dalam menangkap secara akurat pola data historis relatif terhadap konfigurasi parameter alternatif.

b. Model Wilayah Semarang

Tabel 4.3 Model ARIMA Semarang

MODEL ARIMA	**MSE**
0, 1, 1	0,100740
1, 1, 1	0,0911438
1, 1, 0	0,180783

Hasil validasi model ARIMA Kota Semarang disajikan pada Tabel 4.3 dengan menggunakan tiga konfigurasi model yang berbeda: ARIMA (0, 1, 1), ARIMA (1, 1, 1), dan ARIMA (1, 1, 0). *Mean Squared Error* (MSE) digunakan sebagai metrik untuk mengevaluasi kinerja prediktif model. Di antara ketiga konfigurasi tersebut, model ARIMA (1, 1, 1) menunjukkan kinerja tertinggi dengan nilai MSE terendah sebesar 0,0911438, disusul model ARIMA (0, 1, 1) dengan nilai MSE sebesar 0,100740. Sebaliknya model ARIMA (1, 1, 0) menunjukkan nilai MSE tertinggi sebesar 0,180783. Temuan ini menunjukkan bahwa model ARIMA dengan parameter p=1, d=1, dan q=1 paling efektif dalam memprediksi data Kota Semarang karena menghasilkan kesalahan prediksi terkecil. Oleh karena itu, validasi ini menggarisbawahi keunggulan model ARIMA dengan parameter p=1, d=1, dan q=1 dalam menangkap pola data historis di Kota Semarang dibandingkan konfigurasi lainnya, sehingga menjadikannya model yang lebih tepat untuk dianalisis.

c. Model Wilayah Surabaya

Tabel 4.4 Model ARIMA Surabaya

MODEL ARIMA	MSE
0, 1, 1	0,174431
1, 1, 1	0,157882
1, 1, 0	0,263248

Hasil validasi model ARIMA untuk Kota Surabaya disajikan pada Tabel 4.4 yang menampilkan pengujian tiga konfigurasi model ARIMA yang berbeda: ARIMA (0, 1, 1), ARIMA (1, 1, 1), dan ARIMA (1, 1, 0). Metrik *Mean Squared Error* (MSE) digunakan untuk menilai kinerja prediktif setiap model. Temuan menunjukkan bahwa model ARIMA (1, 1, 1) menunjukkan nilai MSE terendah sebesar 0,157882, yang menunjukkan kinerja yang unggul dibandingkan konfigurasi lainnya. Sebaliknya, model ARIMA (0, 1, 1) menghasilkan MSE sebesar 0,174431, dan model ARIMA (1, 1, 0) menunjukkan kinerja terlemah dengan MSE sebesar 0,263248. Oleh karena itu, berdasarkan hasil tersebut dapat disimpulkan bahwa model ARIMA (1, 1, 1) paling akurat untuk peramalan data di Surabaya. Validasi ini menggarisbawahi kemanjuran model dengan parameter p=1, d=1, dan q=1 dalam memberikan prediksi yang lebih tepat dengan pengurangan kesalahan, sehingga menjadikannya pilihan yang cocok untuk analisis data di Surabaya.

4.4 Evaluasi Model

Dalam subab ini, dilakukan tahap evaluasi model ARIMA yang digunakan untuk peramalan di tiga wilayah berbeda yaitu Kabupaten Sleman, Kota Semarang, dan Kota Surabaya. Evaluasi dilakukan dengan menggunakan dua metrik utama, yaitu *Mean Squared Error* (MSE) dan *Mean Absolute Percentage Error* (MAPE). MSE memberikan informasi mengenai besar kesalahan prediksi dalam skala asli data, sedangkan MAPE memberikan informasi mengenai besar kesalahan prediksi dalam bentuk persentase.

4.4.1 *Mean Squared Error* (MSE)

MSE dihitung untuk mengevaluasi performa model dalam setiap wilayah. Nilai MSE yang lebih rendah menunjukkan model yang lebih baik karena memiliki kesalahan prediksi yang lebih kecil. Berikut adalah tabel MSE untuk ketiga wilayah:

Tabel 4.5 MSE Wilayah Penelitian

Wilayah	Model ARIMA	MSE
Kabupaten Sleman	(0, 1, 1)	0,865332
Semarang	(1, 1, 1)	0,0911438
Surabaya	(1, 1, 1)	0,157882

Dari tabel di atas, terlihat bahwa model ARIMA dengan konfigurasi (1, 1, 1) memberikan performa terbaik untuk Kota Semarang dan Kota Surabaya, sedangkan konfigurasi (0, 1, 1) lebih cocok untuk Kabupaten Sleman.

4.4.2*Mean Absolute Percentage Error* (MAPE)

MAPE dihitung untuk mengevaluasi performa model dalam persentase kesalahan. Berikut adalah nilai MAPE untuk ketiga wilayah:

Tabel 4.6 MAPE Wilayah Penelitian

Wilayah	Model ARIMA	MAPE (%)
Kabupaten Sleman	(0, 1, 1)	8,53
Semarang	(1, 1, 1)	9,11
Surabaya	(1, 1, 1)	15,78

Berdasarkan tabel di atas, model ARIMA (0, 1, 1) untuk Kabupaten Sleman memiliki MAPE sebesar 8.53%, model ARIMA (1, 1, 1) untuk Kota Semarang memiliki MAPE sebesar 9.11%, dan model ARIMA (1, 1, 1) untuk Kota Surabaya memiliki MAPE sebesar 15.78%. Semakin kecil nilai MAPE, semakin baik performa model dalam peramalan data.

4.4.3Tabel Range MSE dan MAPE

Untuk memberikan gambaran yang lebih jelas, berikut adalah tabel range untuk MSE dan MAPE:

Tabel 4.7 Range MSE dan MAPE

Range	MSE	MAPE (%)
Sangat Baik	0,00 – 0,10	0,00 – 5,00
Baik	0,11 – 0,50	5,01 – 10,00
Cukup Baik	0,51 – 1,00	10,01 – 20,00
Buruk	> 1,00	> 20,00

Dari hasil evaluasi, berikut adalah kesimpulan:

Kabupaten Sleman: Model ARIMA (0, 1, 1) dengan MSE sebesar 0.865332 dan MAPE sebesar 8.53%. Berdasarkan tabel range, model ini berada dalam kategori "Baik" untuk MAPE dan "Cukup Baik" untuk MSE.

Kota Semarang: Model ARIMA (1, 1, 1) dengan MSE sebesar 0.0911438 dan MAPE sebesar 9.11%. Berdasarkan tabel range, model ini berada dalam kategori "Sangat Baik" untuk MSE dan "Baik" untuk MAPE.

Kota Surabaya: Model ARIMA (1, 1, 1) dengan MSE sebesar 0.157882 dan MAPE sebesar 15.78%. Berdasarkan tabel range, model ini berada dalam kategori "Baik" untuk MSE dan "Cukup Baik" untuk MAPE.

Tabel 4.8 Evaluasi MSE dan MAPE

Wilayah	Model ARIMA Terbaik	MSE	MAPE (%)
Kabupaten Sleman	(0,1,1)	0.865332	8.53%
Semarang	(1,1,1)	0.0911438	9.11%
Surabaya	(1,1,1)	0.157882	15.78%

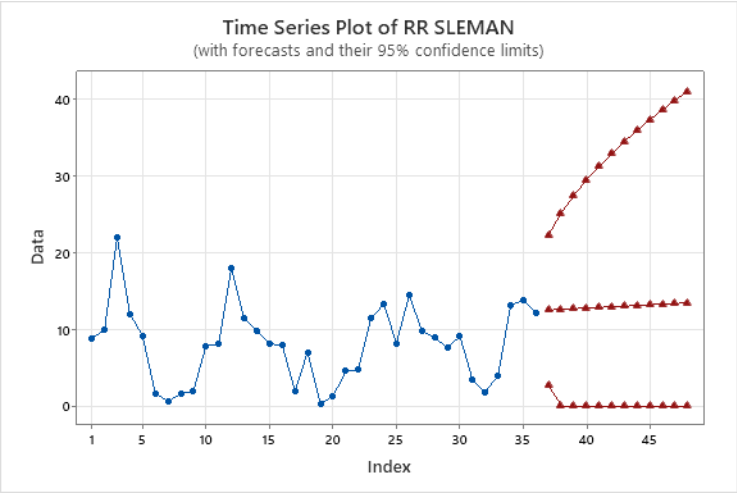
4.4.4Kesimpulan Model yang Cocok

Berdasarkan evaluasi MSE dan MAPE, dapat disimpulkan bahwa:

Untuk Kabupaten Sleman, model ARIMA (0, 1, 1) adalah yang paling cocok, untuk Kota Semarang, model ARIMA (1, 1, 1) adalah yang paling cocok, untuk Kota Surabaya, model ARIMA (1, 1, 1) adalah yang paling cocok.

4.5 Peramalan Curah Hujan

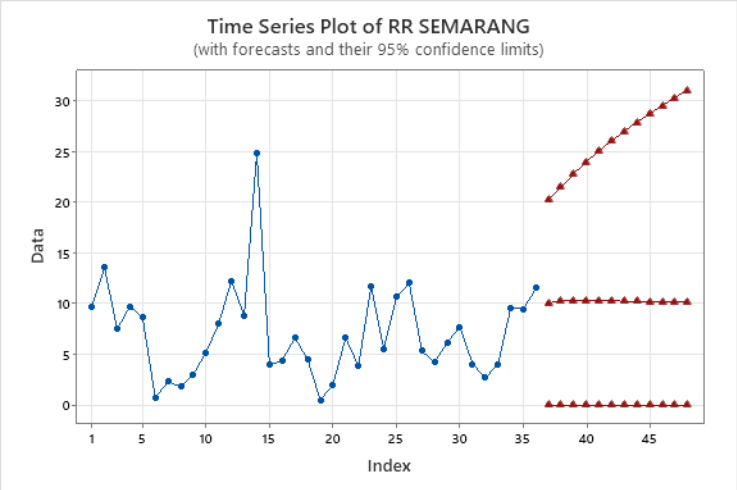
a. Prediksi Curah Hujan Wilayah Kabupaten Sleman



Gambar 4.17 Hasil Plot Time Series Kabupaten Sleman ARIMA (0, 1, 1)

Plot time series Wilayah Kabupaten Sleman menggambarkan variasi nilai Rainfal Rate (RR) dari waktu ke waktu di Sleman. Plot tersebut menunjukkan fluktuasi nilai Rainfal Rate (RR) yang signifikan, terutama pada periode awal dan pertengahan. Prediksi masa depan ditandai dengan tiga garis merah, dengan garis tengah mewakili nilai rata-rata Rainfall Rate (RR) yang diharapkan untuk periode mendatang. Perkiraan nilai Rainfall Rate (RR) adalah sebagai berikut: 12,5 mm pada bulan pertama, sedikit meningkat menjadi 12,6 mm pada bulan kedua, dan selanjutnya menjadi 12,7 mm pada bulan ketiga. Bulan-bulan berikutnya menunjukkan peningkatan nilai prediksi Rainfall Rate (RR) secara bertahap hingga mencapai 13,4 mm pada bulan kesebelas dan kedua belas. Batas atas dan bawah perkiraan, diwakili oleh dua garis merah tambahan, melebar seiring dengan meluasnya perkiraan ke masa depan, yang menunjukkan meningkatnya ketidakpastian dalam prediksi. Ketidakpastian ini penting untuk dipertimbangkan dalam analisis untuk mengantisipasi potensi variasi nilai Rainfall Rate (RR) di masa depan.

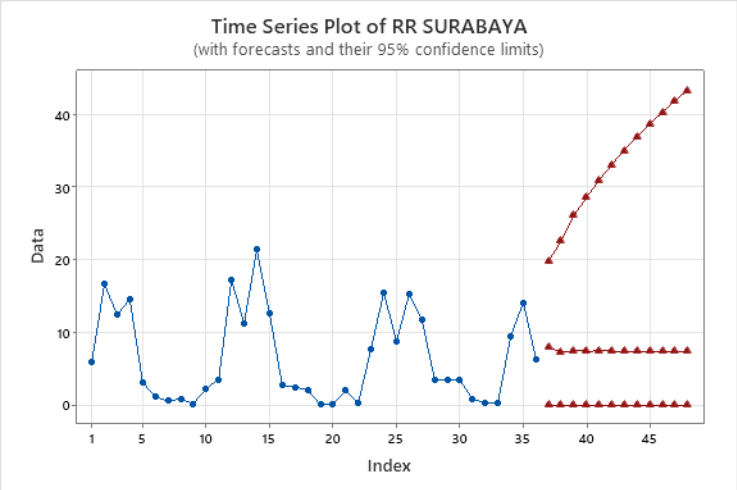
b. Prediksi Curah Hujan Wilayah Semarang



Gambar 4.18 Hasil Plot Time Series Semarang ARIMA (1, 1, 1)

Plot time series Wilayah Semarang menggambarkan pola curah hujan yang berfluktuasi di wilayah tersebut, dengan puncak yang jelas terlihat pada sekitar periode ke-15. Setelah puncak ini, nilai Rainfall Rate (RR) menunjukkan lebih stabil dengan sedikit fluktuasi. Prakiraan curah hujan di masa depan, yang ditandai dengan tiga garis merah, menunjukkan tren menuju stabilitas. Garis merah tengah berfungsi sebagai garis prediksi utama yang memperkirakan nilai Rainfall Rate (RR) rata-rata di masa depan. Proyeksi menunjukkan bahwa curah hujan di Semarang diperkirakan sebesar 10,0 mm pada bulan pertama, sedikit meningkat menjadi 10,3 mm pada bulan kedua dan ketiga, dan mempertahankan tingkat ini hingga bulan keenam. Sedikit penurunan menjadi 10,2 mm diperkirakan terjadi pada bulan ketujuh, dan tetap konsisten hingga bulan kedua belas. Garis merah atas dan bawah mewakili batas perkiraan yang menunjukkan tingkat ketidakpastian dalam prediksi. Batasan ini semakin melebar seiring berjalannya waktu, mencerminkan meningkatnya ketidakpastian. Memahami batas-batas ini sangat penting untuk perencanaan dan pengambilan keputusan yang efektif, sehingga memberikan wawasan tentang potensi variasi yang mungkin terjadi pada pola curah hujan di masa depan.

c. Prediksi Curah Hujan Wilayah Surabaya



Gambar 4.19 Hasil Plot Time Series Surabaya ARIMA (1, 1, 1)

Plot time series yang disajikan menggambarkan pola curah hujan di Wilayah Surabaya. Nilai Rainfall Rate (RR) di Surabaya terlihat menunjukkan tren yang relatif stabil dibandingkan lokasi lain, dengan sedikit fluktuasi yang terjadi pada periode tertentu. Plotnya juga mencakup prediksi masa depan yang diwakili oleh tiga garis merah, dengan garis tengah berfungsi sebagai prediksi utama yang memperkirakan nilai Rainfall Rate (RR) rata-rata. Proyeksi nilai Rainfall Rate (RR) untuk Surabaya adalah sebagai berikut: 7,9 mm pada bulan pertama, turun menjadi 7,3 mm pada bulan kedua, selanjutnya meningkat menjadi 7,5 mm pada bulan ketiga, dan stabil pada kisaran 7,4 mm pada bulan keempat hingga kedua belas. bulan. Garis batas atas dan bawah yang ditampilkan dalam plot menunjukkan kisaran penurunan nilai Rainfall Rate (RR) aktual, yang mencerminkan meningkatnya ketidakpastian dalam prediksi seiring berjalannya waktu. Oleh karena itu, meskipun garis prediksi utama memberikan wawasan tentang nilai Rainfall Rate (RR) yang diharapkan, batas atas dan bawah menekankan pentingnya mempertimbangkan ketidakpastian dalam analisis dan perencanaan selanjutnya.

4.5.1Kesimpulan Peramalan Curah Hujan

Berdasarkan hasil prediksi yang telah didapatkan, terdapat garis merah pada plot time series menggarisbawahi pentingnya memperhatikan variasi geografis dalam mempelajari pola cuaca di Pulau Jawa.

Tabel 4.9 Kategori Curah Hujan Bulanan

Curah Hujan (MM/Bulan)	**Kategori**
0 – 100	Rendah
100 – 300	Menengah
300 – 500	Tinggi
>500	Sangat Tinggi

Sumber: BMKG

Pemilihan Kabupaten Sleman, Semarang, dan Surabaya sebagai objek studi sangat tepat karena setiap kota memiliki karakteristik geografis yang berbeda, yang tercermin dalam variasi curah hujan mereka. Di Kabupaten Sleman, prediksi menunjukkan fluktuasi curah hujan yang signifikan, sementara Semarang menunjukkan stabilitas setelah puncak tertentu, dan Surabaya memperlihatkan tren yang relatif stabil. Analisis ini menunjukkan bagaimana faktor geografis berperan penting dalam mempengaruhi pola curah hujan, yang penting dalam perencanaan sumber daya dan mitigasi bencana.

Dengan prediksi yang menunjukkan perbedaan signifikan di antara ketiga kota ini, analisis ini memberikan wawasan mendalam tentang potensi tren cuaca di masa depan. Oleh karena itu, prediksi curah hujan ini tidak hanya mendukung pemilihan kota-kota tersebut sebagai objek studi tetapi juga memberikan pemahaman yang lebih baik mengenai bagaimana perbedaan geografis dapat mempengaruhi pola cuaca, yang penting untuk analisis lebih lanjut dan pengambilan keputusan.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil menerapkan metode *Autoregressive Integrated Moving Average* (ARIMA) untuk memprediksi curah hujan di Kabupaten Sleman, Semarang, dan Surabaya selama periode 2020 hingga 2022. Hasil analisis menunjukkan bahwa model ARIMA (0, 1, 1) memberikan hasil prediksi terbaik untuk Kabupaten Sleman dengan *Mean Squared Error* (MSE) sebesar 0,865332, meskipun tidak optimal. Kota Semarang menunjukkan kinerja paling baik dengan model ARIMA (1, 1, 1) yang memiliki MSE terendah sebesar 0,0911438, menandakan akurasi prediksi yang sangat baik. Surabaya juga menggunakan model ARIMA (1, 1, 1) dengan MSE sebesar 0,157882, menunjukkan performa baik meskipun terdapat ketidakpastian yang lebih tinggi dibandingkan Semarang. Evaluasi kinerja model melalui nilai *Mean Absolute Percentage Error* (MAPE) menunjukkan bahwa Kabupaten Sleman, Semarang, dan Surabaya masing-masing memiliki MAPE sebesar 8,53%, 9,11%, dan 15,78%, dengan semakin rendahnya nilai MAPE menunjukkan semakin baiknya kinerja model dalam memprediksi curah hujan. Perbedaan karakteristik geografis dan iklim antara ketiga wilayah mencerminkan variasi dalam pola curah hujan, di mana Sleman mengalami fluktuasi yang signifikan, Semarang menunjukkan stabilitas setelah puncak tertentu, dan Surabaya memiliki tren yang relatif stabil.

5.2 Saran

Berdasarkan temuan penelitian, beberapa saran yang dapat diajukan adalah sebagai berikut. Pertama, diperlukan pengumpulan data yang lebih lengkap dan berkelanjutan untuk meningkatkan akurasi model, serta mempertimbangkan penggunaan metode prediksi lain, seperti model *hybrid* yang menggabungkan ARIMA dengan teknik machine learning. Kedua, penelitian lanjutan perlu dilakukan untuk mengeksplorasi dampak perubahan iklim terhadap pola curah hujan di wilayah lain di Indonesia. Penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap pemahaman dan pengelolaan curah hujan di sektor yang rentan terhadap perubahan cuaca.

DAFTAR PUSTAKA

Al'afi, A. M., Widiarti, Kurniasari, D., & Usman, M. (2020). Peramalan Data Time Series Seasonal Menggunakan Metode Analisis Spektral Berdasarkan data yang tersedia diperoleh model terbaik untuk peramalan penumpang pesawat di Bandar Udara Raden Intan II adalah Seasonal ARIMA (0. In *Jurnal Siger Matematika* (Vol. 01, Issue 01).

Alexander, & Harahab, S. (2009). *LAPORAN TUGAS AKHIR PERENCANAAN EMBUNG TAMBABOYO KABUPATEN*.

Anshari, M. K., Arifin, S., & Rahmadiansah, A. (2013). Perancangan Prediktor Cuaca Maritim Berbasis Logika Fuzzy Menggunakan User Interface Android. *Jurnal Teknik ITS, 2(2), A324-A328.*

Aulia, N. N., Gunawan, P. H., & Rohmawati, A. A. (2018). Prediksi Curah Hujan Menggunakan Gerak Brown dan Rataan Tahunan Data Pada Missing Values. *Indonesian Journal on Computing (Indo-JC)*, *3*(2), 71. https://doi.org/10.21108/indojc.2018.3.2.233

Breiman, L. (2001). *Random Forests* (Vol. 45).

Box, G.E.P. & Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control, Revised Edition, San Francisco: Holden Day.

Burrahman, H., Silitonga, A. K., Batubara, I. H., & Fadlan, A. (2018). *Prosiding SNFA (Seminar Nasional Fisika dan Aplikasinya) 2018 PENGARUH ASIMILASI MODEL CUACA WEATHER RESEARCH FORECAST (WRF) DENGAN DATA RADIASI SATELIT TERHADAP ESTIMASI CURAH HUJAN (Studi Kasus Stasiun Meteorologi Pattimura-Ambon Tanggal 24-25 Juli 2013)*.

Fadlan, A., Safril, A., Suwandi, Veanti, D. P. O., Nugraheni, I. R., Septiadi, D., Harahap, D., Nuraini, N., & Munawar. (2022). *Pengetahuan Tentang Iklim dan Cuaca Untuk Kemajuan Pertanian di Kabupaten Indramayu Jawa Barat*.

Fatimah, D. D. S., & Rahmawati, E. (2022). *Penggunaan Metode Decision Tree dalam Rancang Bangun Sistem Prediksi untuk Kelulusan Mahasiswa*. https://jurnal.itg.ac.id/

Haghshenas, S. S., Pirouz, B., Haghshenas, S. S., Pirouz, B., Piro, P., Na, K. S., Cho, S. E., & Geem, Z. W. (2020). Prioritizing and analyzing the role of climate and urban parameters in the confirmed cases of COVID-19 based on artificial intelligence applications. *International Journal of Environmental Research and Public Health*, *17*(10). https://doi.org/10.3390/ijerph17103730

Herlina, N., & Prasetyorini, A. (2020). Pengaruh Perubahan Iklim pada Musim Tanam dan Produktivitas Jagung (Zea mays L.) di Kabupaten Malang (Effect of Climate Change on Planting Season and Productivity of Maize (Zea mays L.) in Malang Regency). *Jurnal Ilmu Pertanian Indonesia (JIPI)*, *Januari*, *25*(1), 118–128. https://doi.org/10.18343/jipi.25.1.118

Hutasuhut, A. H., Anggraeni, W., & Tyasnurita, R. (2014). Pembuatan aplikasi pendukung keputusan untuk peramalan persediaan bahan baku produksi plastik blowing dan inject menggunakan metode ARIMA (Autoregressive Integrated Moving Average) di CV. Asia. *Jurnal Teknik ITS, 3(2), A169-A174.*

Khan, M. Y., Qayoom, A., Nizami, M. S., Siddiqui, M. S., Wasi, S., & Raazi, S. M. K. U. R. (2021). Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques. *Complexity*, *2021*. https://doi.org/10.1155/2021/2553199

Kim, R.-H., Lee, S., Jeong, J., Lee, J.-H., & Kim, Y.-K. (2007). Reuse of greywater and rainwater using fiber filter media and metal membrane. *Desalination*, *202*(1–3), 326–332. https://doi.org/10.1016/j.desal.2005.12.071

Laia, M. L., & Setyawan, Y. (2020). Perbandingan hasil klasifikasi curah hujan menggunakan metode SVM dan NBC. *Jurnal Statistika Industri Dan Komputasi, 5(02), 51-61.*

Masum, S. J. H. (2019). *Climatic Hazards in Bangladesh-A Literature Review*. https://doi.org/10.13140/RG.2.2.10239.94882

Murphy, A. H., & Brown, B. G. (1984). *A Comparative Evaluation of Objective and Subjective Weather Forecasts in the United States Weather forecasting/forecasts Objective and subjective fore-casting/forecasts Probability forecasting/forecasts Evaluation of forecasts Reliability, accuracy and skill of forecasts Trends in forecast performance Relationship between objective and subjective forecasts Weather fore-casting and forecasting in other fields.*

Nifa, K., Boudhar, A., Ouatiki, H., Elyoussfi, H., Bargam, B., & Chehbouni, A. (2023). Deep Learning Approach with LSTM for Daily Streamflow Prediction in a Semi-Arid Area: A Case Study of Oum Er-Rbia River Basin, Morocco. *Water (Switzerland)*, *15*(2). https://doi.org/10.3390/w15020262

Nur, I., & Astuti, S. P. (2006). Mengolah data statistik dengan mudah menggunakan minitab 14. *Yogyakarta: Andi.*

Polawan, S. S. M., & Alam, F. (2019). *MEMAHAMI BENCANA BANJIR DAN LONGSOR*. https://www.researchgate.net/publication/346678807

Rokach, L., & Maimon, O. (2006). Decision Trees. In *Data Mining and Knowledge Discovery Handbook* (pp. 165–192). Springer-Verlag. https://doi.org/10.1007/0-387-25465-x_9

Safar, N. Z. M., Ramli, A. A., Mahdin, H., Ndzi, D., & Khalif, K. M. N. K. (2019). Rain prediction using fuzzy rule based system in North-West Malaysia. *Indonesian Journal of Electrical Engineering and Computer Science*, *14*(3), 1572–1581. https://doi.org/10.11591/ijeecs.v14.i3.pp1572-1581

Salman, A. G., Heryadi, Y., Abdurahman, E., & Suparta, W. (2018). Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting. *Procedia Computer Science*, *135*, 89–98. https://doi.org/10.1016/j.procs.2018.08.153

Shumway, R. H., & Stoffer, D. S. (2000). Time series regression and ARIMA models. *Time Series Analysis and Its Applications, 89-212.* https://doi.org/https://doi.org/10.1007/978-1-4757-3261-0_2

Singh, N., Chaturvedi, S., & Akhter, S. (2019). *2019 International Conference on Signal Processing and Communication (ICSC)*. IEEE.

Tyralis, H., & Papacharalampous, G. (2017). Variable selection in time series forecasting using random forests. *Algorithms*, *10*(4). https://doi.org/10.3390/a10040114

Watta, P. B., & Hassoun, M. H. (1995). *Experiments in Machine Learning Using Artificial Neural Networks for Control and Image Compression*.

Wulandari, R. A., & Gernowo, R. (2019). *METODE AUTOREGRESSIVE INTEGRATED MOVINGAVERAGE (ARIMA) DAN METODE ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS) DALAM ANALISIS CURAH HUJAN* (Vol. 22, Issue 1).

LAMPIRAN

https://drive.google.com/drive/folders/1PKrGYlZ_jeZ1L2IwaCKcf42wMa6oHIVI?usp=sharing