

PREDIKSI *URBAN HEAT ISLAND* MENGGUNAKAN METODE *RANDOM FOREST REGRESSION* (STUDI KASUS : KOTA MALANG)

SKRIPSI

Oleh:
ILHAM SHODIQ MU'ARIF BILLAH
NIM. 200605110072



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**PREDIKSI *URBAN HEAT ISLAND* MENGGUNAKAN METODE
RANDOM FOREST REGRESSION (STUDI KASUS : KOTA MALANG)**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
ILHAM SHODIQ MU'ARIF BILLAH
NIM. 200605110072

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PERSETUJUAN

PREDIKSI *URBAN HEAT ISLAND* MENGGUNAKAN METODE
RANDOM FOREST REGRESSION (STUDI KASUS : KOTA MALANG)

SKRIPSI

Oleh:
ILHAM SHODIQ MU'ARIF BILLAH
NIM. 200605110072

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 20 September 2024

Pembimbing I,



Dr. Yunifa Miftachul Arif, M.T
NIP. 19830616 201101 1 004


Pembimbing II,



Hani Nurhayati, M.T
NIP. 19780625 200801 2 006

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Ir. Pachi Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

**PREDIKSI URBAN HEAT ISLAND MENGGUNAKAN METODE
RANDOM FOREST REGRESSION (STUDI KASUS : KOTA MALANG)**

SKRIPSI

Oleh:
ILHAM SHODIQ MU'ARIF BILLAH
NIM. 200605110072

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 2 Oktober 2024

Susunan Dewan Penguji

Ketua Penguji : Dr. Fresy Nugroho, M.T
NIP. 19710722 201101 1 001

Anggota Penguji I : Okta Qomaruddin Aziz, M. Kom
NIP. 19911019 201903 1 013

Anggota Penguji II : Dr. Yunifa Miftachul Arif, M.T
NIP. 19830616 201101 1 004

Anggota Penguji III : Hani Nurhayati, M.T
NIP. 19780625 200801 2 006



Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachrul Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Ilham Shodiq Mu'arif Billah
NIM : 200605110072
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : PREDIKSI URBAN HEAT ISLAND
MENGUNAKAN METODE RANDOM
FOREST REGRESSION (STUDI KASUS :
KOTA MALANG)

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 2 Oktober 2024
Yang membuat pernyataan,



ILHAM SHODIQ M. B.
NIM. 200605110072

MOTTO

“To Be Slave of Allah”

HALAMAN PERSEMBAHAN

Puji Syukur atas kehadiran Allah Subhanahu wa ta'ala, karena berkat rahmat dan petunjuk-Nya penulis dapat menyelesaikan skripsi ini. Shalawat serta salam kepada Rasulullah Shallallahu 'alaihi wasallam, yang telah membawa kita dari zaman *jahiliyah* menuju *addinul Islam*.

Penulis mempersembahkan karya ini kepada kedua orang tua, kerabat, teman, kakak tingkat, dan sahabat yang telah menemani perkuliahan daring maupun luring. Tanpa bantuan mereka penulis tidak akan bisa sampai di titik ini.

KATA PENGANTAR

Assalamualaikum wr wb.

Puji syukur penulis panjatkan kepada kehadiran Allah Swt yang telah melimpahkan nikmat serta karunia-Nya sehingga penulis mampu menyelesaikan penulisan Skripsi yang berjudul “Prediksi Urban Heat Island Menggunakan Metode *Random Forest Regression* (Studi Kasus : Kota Malang)” dengan baik.

Dalam penulisan skripsi ini, penulis menyadari banyak pihak yang terlibat baik dalam proses membimbing penulisan dan juga memberikan semangat dan dukungan moril atau materil. Untuk itu, penulis ingin menyampaikan terima kasih kepada:

1. Prof. Dr. H.M. Zainuddin, MA., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT, IPU, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Yunifa Miftachul Arif, M.T selaku Dosen Pembimbing I yang telah memberikan banyak dukungan dan bimbingan dalam penulisan skripsi ini.
5. Dr. Fresy Nugroho, M.T. selaku Ketua Penguji dan Okta Qomaruddin Aziz, M.Kom. selaku dosen penguji I yang telah menguji serta memberikan masukan dalam penulisan skripsi ini.

6. Seluruh Dosen dan Jajaran Staf Program Studi Teknik Informatika yang telah memberikan banyak bantuan dalam skripsi ini.
7. Kedua orang tua, Bapak Irfan Mu'tasim Billah dan Ibu Umi Mutaqiyatul Azizah, S.Ag. Kemudian adik Farhan Nidzom Amali serta kedua adik kecil saya Farhana Arizkiyah dan Abdullah yang telah mendahului saya yang insyaallah sudah berada di Surga-Nya yang telah memberikan dukungan kepada penulis untuk menyelesaikan skripsi ini.
8. Teman-teman dari "INGFO LYFE" dan "WARGA KOS 61A" yaitu Tegar, Aji, Adam, Fania, Rani, Rini(Bela), Aita, Alfina, dan Daniyal yang telah menemani dan membantu saya saat dalam masa-masa terpuruk sehingga saya dapat bangkit untuk melanjutkan skripsi ini.
9. Teman-teman Angkatan 2020 Teknik Informatika "INTEGER" yang telah memberikan dukungan, semangat, dan kebersamaan selama masa studi dan penyusunan skripsi ini.

Akhir kata, penulis mengakui bahwa penulisan pada skripsi ini masih banyak kekurangan.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 2 Oktober 2024

Penulis

DAFTAR ISI

| | |
|--|-----------|
| HALAMAN PENGAJUAN..... | ii |
| HALAMAN PERSETUJUAN | iii |
| HALAMAN PENGESAHAN..... | iv |
| PERNYATAAN KEASLIAN TULISAN | v |
| MOTTO | vi |
| HALAMAN PERSEMBAHAN | vii |
| KATA PENGANTAR..... | viii |
| DAFTAR ISI..... | x |
| DAFTAR TABEL | xii |
| DAFTAR GAMBAR..... | xiii |
| ABSTRAK | xiv |
| ABSTRACT..... | xv |
| مستخلص البحث | xvi |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 5 |
| 1.3 Batasan Masalah..... | 5 |
| 1.4 Tujuan Penelitian | 5 |
| 1.5 Manfaat Penelitian | 5 |
| BAB II KAJIAN PUSTAKA..... | 6 |
| 2.1 Penelitian Terdahulu | 6 |
| 2.2 <i>Urban Heat Island (UHI)</i> | 8 |
| 2.3 <i>Land Surface Temperature(LST)</i> | 11 |
| 2.4 <i>Random Forest</i> | 12 |
| 2.5 Evaluasi Performa | 14 |
| 2.5.1 <i>Mean Square Error (MSE)</i> | 14 |
| 2.5.2 <i>Mean Absolute Percentage Error (MAPE)</i> | 15 |
| 2.5.3 <i>R-Squared (R²)</i> | 15 |
| BAB III DESAIN DAN IMPLEMENTASI | 17 |
| 3.1 Desain Sistem..... | 17 |
| 3.2 <i>Data Collection</i> | 17 |
| 3.2.1 Satelit Landsat-8 | 18 |
| 3.2.2 Satelit Sentinel-2 | 18 |
| 3.2.3 Satelit SRTM..... | 19 |
| 3.3 <i>Data Preparation</i> | 20 |
| 3.3.1 <i>Cloud Masking</i> | 20 |
| 3.3.2 <i>Features Calculation</i> | 20 |
| 3.4 <i>Data Preprocessing</i> | 27 |

| | |
|---|-----------|
| 3.4.1 <i>Data Correlation</i> | 28 |
| 3.4.2 Normalisasi | 29 |
| 3.5 <i>Modeling</i> | 32 |
| 3.5.1 <i>Splitting Data</i> | 32 |
| 3.5.2 <i>Training Model</i> | 32 |
| 3.6 <i>Evaluation Performance</i> | 38 |
| 3.7 Skenario..... | 38 |
| BAB IV UJI COBA DAN PEMBAHASAN | 40 |
| 4.1 Hasil Uji Coba..... | 40 |
| 4.1.1 Hasil Uji Coba Skenario 1..... | 40 |
| 4.1.2 Hasil Uji Coba Skenario 2..... | 49 |
| 4.2 Analisis dan Pembahasan..... | 58 |
| 4.3 Perbandingan Hasil Visualisasi..... | 69 |
| 4.4 Integrasi dalam Islam | 72 |
| 4.4.1 Muamalah Ma'a Allah | 72 |
| 4.4.2 Muamalah Ma'an-Nas..... | 74 |
| 4.4.3 Muamalah Ma'a Alam | 76 |
| BAB V KESIMPULAN DAN SARAN | 79 |
| 5.1 Kesimpulan | 79 |
| 5.2 Saran..... | 80 |
| DAFTAR PUSTAKA | |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1 Penelitian Terkait..... | 7 |
| Tabel 2.2 Nilai MAPE | 15 |
| Tabel 3.1 <i>Dataset Features</i> | 20 |
| Tabel 3.2 Data sampel contoh hasil normalisasi..... | 31 |
| Tabel 3.3 Data sampel untuk <i>bootstrap sampling</i> | 33 |
| Tabel 3.4 Contoh hasil <i>bootstrap sampling</i> | 34 |
| Tabel 3.5 Subset 1 node kiri..... | 35 |
| Tabel 3.6 Subset 2 node kanan..... | 35 |
| Tabel 3.7 Skenario uji..... | 38 |
| Tabel 4.1 Hasil skenario 1 agregasi mean..... | 40 |
| Tabel 4.2 Hasil skenario 2 agregasi median | 49 |
| Tabel 4.3 Rata-rata MAPE parameter <i>ratio</i> | 59 |
| Tabel 4.4 Rata-rata MAPE parameter <i>trees</i> | 61 |
| Tabel 4.5 Rata-rata MAPE parameter <i>Max Depth Trees</i> | 63 |
| Tabel 4.6 Rata-rata MAPE parameter <i>Bootstrap Samples</i> | 65 |
| Tabel 4.7 Rata-rata evaluasi skenario 1 dan skenario 2..... | 68 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1 Ilustrasi efek dari Urban Heat Island Sumber: (T & VK, 2018) | 9 |
| Gambar 2.2 <i>Random Forest</i> Sumber: builtin.com | 13 |
| Gambar 3.1 Desain sistem dan penelitian..... | 17 |
| Gambar 3.2 NDVI..... | 22 |
| Gambar 3.3 NDBI..... | 23 |
| Gambar 3.4 NDWI..... | 24 |
| Gambar 3.5 UDI..... | 27 |
| Gambar 3.6 <i>Data Correlation</i> | 29 |
| Gambar 3.7 Data sebelum normalisasi | 30 |
| Gambar 3.8 Data setelah normalisasi..... | 31 |
| Gambar 4.1 Evaluasi MSE skenario 1 agregasi mean | 44 |
| Gambar 4.2 Evaluasi MAPE skenario 1 agregasi mean | 46 |
| Gambar 4.3 Evaluasi R2 skenario 1 agregasi mean..... | 48 |
| Gambar 4.4 Evaluasi MSE skenario 2 agregasi median | 53 |
| Gambar 4.5 Evaluasi MAPE skenario 2 agregasi median | 55 |
| Gambar 4.6 Evaluasi R2 skenario 2 agregasi median..... | 57 |
| Gambar 4. 7 Visualisasi rata-rata MAPE pada parameter ratio | 59 |
| Gambar 4.8 Visualisasi rata-rata MAPE pada parameter <i>trees</i> | 62 |
| Gambar 4.9 Visualisasi rata-rata MAPE pada parameter <i>Max Depth Trees</i> | 64 |
| Gambar 4.10 Visualisasi rata-rata MAPE pada parameter <i>Bootstrap Samples</i> | 66 |
| Gambar 4.11 Visualisasi perbandingan rata-rata evaluasi skenario 1 dan skenario 2 | 68 |
| Gambar 4.12 Visualisasi citra perbandingan LST aktual dan prediksi 2019-2023..... | 70 |
| Gambar 4.13 Analisa persebaran <i>high error</i> prediksi UHI 2019-2023..... | 70 |

ABSTRAK

Billah, Ilham Shodiq Mu'arif. 2024. **Prediksi *Urban Heat Island* Menggunakan Metode *Random Forest Regression* (Studi Kasus: Kota Malang)**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Yunifa Miftachul Arif, M.T (II) Hani Nurhayati. M.T.

Kata kunci: Prediksi, *Urban Heat Island*, *Land Surface Temperature*, *Random Forest Regression*

Pada abad ke-21 ini *Urban Heat Island* atau pulau panas perkotaan merupakan salah satu masalah besar yang sangat berdampak pada peradaban umat manusia. Salah satu kota besar yang terdampak ini adalah Kota Malang. Penelitian ini bertujuan untuk merancang sistem prediksi *Urban Heat Island* menggunakan metode *Random Forest Regression* dengan bertujuan untuk mengevaluasi performa model untuk mendapatkan model terbaik. Penelitian ini melibatkan dua skenario utama, yaitu agregasi akhir menggunakan *mean* dan *median*, yang bertujuan untuk menentukan model dengan performa terbaik dalam memprediksi UHI. Evaluasi model dilakukan menggunakan tiga metrik utama: *Mean Squared Error* (MSE), *Mean Absolute Percentage Error* (MAPE), dan *R-squared* (R^2). Hasil penelitian menunjukkan bahwa model dengan agregasi *mean* memiliki keunggulan dalam hal akurasi prediksi, dengan nilai MSE dan MAPE yang lebih rendah serta nilai R^2 yang lebih tinggi dibandingkan agregasi *median*. Model terbaik ditemukan pada skenario agregasi *mean* dengan konfigurasi 500 *trees*, *max depth* 20, dan *bootstrap samples* 0.5, yang menghasilkan nilai MSE sebesar 1.56068, MAPE sebesar 2.700732, dan R^2 sebesar 0.857488. Selain itu, penerapan model terbaik ini dalam memprediksi UHI per tahun dari 2019 hingga 2023 menunjukkan adanya penurunan area dengan prediksi akurat, yang dapat mencerminkan perubahan lingkungan dan perkembangan urbanisasi di Kota Malang. Berdasarkan temuan ini, model agregasi *mean* dipilih sebagai model terbaik untuk prediksi *Urban Heat Island*.

ABSTRACT

Billah, Ilham Shodiq Mu'arif. 2024. **Prediksi *Urban Heat Island* Menggunakan Metode *Random Forest Regression* (Studi Kasus: Kota Malang)**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Yunifa Miftachul Arif, M.T (II) Hani Nurhayati. M.T.

In the 21st century, Urban Heat Island (UHI) has emerged as a significant issue impacting human civilization. One of the major cities affected by this phenomenon is Malang, Indonesia. This study aims to design a UHI prediction system using the Random Forest Regression method to evaluate model performance and identify the best-performing model. The research involves two main scenarios, utilizing mean and median aggregation, to determine the model with the best performance in predicting UHI. Model evaluation was conducted using three primary metrics: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2). The results indicate that the model using mean aggregation demonstrated an advantage in prediction accuracy, with consistently lower MSE and MAPE values and higher R^2 values compared to the median aggregation. The optimal model was found in the mean aggregation scenario, with a configuration of 500 trees, a max depth of 20, and bootstrap samples set to 0.5, yielding an MSE of 1.56068, MAPE of 2.700732, and R^2 of 0.857488. Furthermore, applying this optimal model to predict UHI annually from 2019 to 2023 showed a gradual decrease in the area with accurate predictions, potentially reflecting environmental changes and urbanization developments in Malang. Based on these findings, the mean aggregation model is selected as the best approach for Urban Heat Island prediction.

Keywords: Prediction, Urban Heat Island, Land Surface Temperature, Random Forest Regression

مستخلص البحث

بِالله، الهام صاديق معريف. 2024. توقع جزر الحرارة الحضرية باستخدام طريقة الانحدار العشوائي للغابات (دراسة حالة: مدينة مالانج). بحث التخرج. برنامج دراسة تقنية المعلومات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (الأول) يونيفا مفتاح العارف، ماجستير يف تكنولوجيا. (الثاني) هاني نور حياتي، ماجستير يف تكنولوجيا. املشرف الثاني

الكلمات المفتاحية: التوقع، جزر الحرارة الحضرية، درجة حرارة سطح الأرض، الانحدار العشوائي للغابات

في القرن الواحد والعشرين، تُعدّ ظاهرة "الجزيرة الحرارية الحضرية" من أكبر المشكلات التي تؤثر بشكل كبير على حضارة البشرية. ومن المدن الكبرى المتأثرة بهذه الظاهرة مدينة مالانج. تهدف هذه الدراسة إلى تصميم نظام للتنبؤ بـ "الجزيرة الحرارية الحضرية" باستخدام طريقة التراجع العشوائي للغابة (*Random Forest Regression*) لتقييم أداء النموذج من أجل الوصول إلى النموذج الأفضل. وتشمل هذه الدراسة سيناريوهين رئيسيين: التجميع باستخدام الوسط الحسابي والتجميع باستخدام الوسيط، بهدف اختيار النموذج الأكثر دقة في التنبؤ بالجزيرة الحرارية الحضرية. تم تقييم النموذج باستخدام ثلاث معايير أساسية: متوسط مربع الخطأ (MSE)، ومتوسط نسبة الخطأ المطلق ($MAPE$)، ومعامل التحديد (R^2) أشارت نتائج الدراسة إلى تفوق النموذج الذي يعتمد على تجميع الوسط الحسابي من حيث دقة التنبؤ، حيث كانت قيم متوسط مربع الخطأ ونسبة الخطأ المطلقة أقل، في حين كانت قيمة معامل التحديد أعلى مقارنة بالتجميع بالوسيط. وقد تم تحديد النموذج الأفضل في سيناريو التجميع بالوسط الحسابي بتهئية شملت 500 شجرة، وعمق أقصى يبلغ 20، وعينات باستخدام "التشغيل المدمج" بنسبة 0.5، حيث حقق هذا النموذج قيمة MSE قدرها 1.56068، و $MAPE$ قدرها 2.700732، و R^2 بلغ 0.857488. بالإضافة إلى ذلك، أظهر تطبيق النموذج الأفضل للتنبؤ بالجزيرة الحرارية الحضرية بين الأعوام 2019 و 2023 انخفاضاً في مساحة المناطق ذات التنبؤ الدقيق، مما قد يعكس التغيرات البيئية والتطور العمراني في مدينة مالانج. وبناءً على هذه النتائج، تم اختيار نموذج التجميع بالوسط الحسابي كأفضل نموذج للتنبؤ بالجزيرة الحرارية الحضرية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada abad ke-21 ini *Urban Heat Island* atau pulau panas perkotaan merupakan salah satu masalah besar yang sangat berdampak pada peradaban umat manusia sebagai akibat dari urbanisasi dan kemajuan industri yang berkembang begitu pesat (RIZWAN et al., 2008). Perkembangan urbanisasi dan industri sendiri tentunya memiliki tujuan yang ingin memajukan peradaban manusia, tetapi ini menjadi masalah dikarenakan ketika mulai banyak manusia yang pindah ke daerah perkotaan dari pedesaan maka lahan pada daerah perkotaan akan menjadi semakin sempit dikarenakan semakin banyaknya populasi di suatu kota. Diperkirakan 50% penduduk dunia (sekitar 3,4 Miliar orang) saat ini tinggal di daerah perkotaan dan angkanya akan terus meningkat setiap tahun dan diperkirakan akan mencapai 60% dari populasi dunia akan pindah ke daerah perkotaan pada tahun 2030 (Mirzaei & Haghghat, 2010). Semakin banyaknya manusia maka akan terjadi kelonjakan permintaan tempat tinggal dan karena itu akan terjadi pembangunan besar-besaran di perkotaan. Dikarenakan kurangnya perencanaan dalam pembangunan ini maka kedepannya akan berdampak buruk terhadap lingkungan.

Dalam konteks studi kasus penelitian ini, yaitu Kota Malang, terdapat sejumlah faktor yang memperkuat urgensi pemahaman terhadap *Urban Heat Island* (UHI). Kota Malang, sebagai salah satu kota besar di Indonesia, telah mengalami pertumbuhan perkotaan yang pesat dalam beberapa dekade terakhir. Pertumbuhan ini tidak hanya mencakup peningkatan jumlah penduduk, tetapi juga ekspansi

wilayah perkotaan, pembangunan infrastruktur, dan aktivitas industri yang semakin intensif. Semua faktor ini berpotensi meningkatkan suhu permukaan dan menciptakan pola urbanisasi yang menghasilkan *Urban Heat Island* (Larasati et al., 2022). Sebagai kota yang memiliki keragaman geografis, Kota Malang juga memiliki kondisi topografi yang beragam, mulai dari dataran rendah hingga daerah perbukitan. Kondisi topografi yang kompleks ini dapat memengaruhi distribusi suhu permukaan di kota, menciptakan pola yang berbeda-beda tergantung pada lokasi geografisnya. Oleh karena itu, pemahaman terhadap fenomena *Urban Heat Island* di Kota Malang menjadi semakin penting untuk mengidentifikasi area-area yang rentan terhadap dampak *Urban heat Island* dan merancang strategi mitigasi yang sesuai.

Pentingnya untuk memperhatikan bahwa pembangunan harus dilakukan dengan cermat agar tidak merusak keseimbangan kondisi alam. Dalam hal ini, aspek lingkungan dan keberlanjutan menjadi faktor utama yang harus dipertimbangkan dalam setiap kebijakan pembangunan kota. Pembangunan yang tidak terencana dengan baik dapat menyebabkan kerusakan lingkungan, penurunan kualitas udara, dan akan berdampak menjadi perubahan iklim. Telah disampaikan dalam Al-Qur'an Surah Al-A'raf ayat 74 yang berbunyi :

وَاذْكُرُوا إِذْ جَعَلْنَاكُمْ خُلَفَاءَ مِنْ بَعْدِ عَادٍ وَبَوَّأْنَاكُمْ فِي الْأَرْضِ تَتَّخِذُونَ مِنْ سُهُولِهَا قُصُورًا وَتَنْحِتُونَ
الْجِبَالَ بَيْوتًا فَاذْكُرُوا الْآءَ اللَّهِ وَلَا تَعْنُوا فِي الْأَرْضِ مُفْسِدِينَ ۚ ٧٤

"Ingatlah ketika (Allah) menjadikan kamu pengganti-pengganti (yang berkuasa) sesudah 'Ad dan memberikan tempat bagimu di bumi. Kamu membuat pada dataran rendahnya bangunan-bangunan besar dan kamu pahat gunung-gunungnya menjadi rumah. Maka, ingatlah nikmat-nikmat Allah dan janganlah kamu melakukan kejahatan di bumi dengan berbuat kerusakan." (QS. Al-A'raf: 74)

Ayat ini mengingatkan kita akan nikmat yang diberikan Allah berupa tempat tinggal dan kemampuan untuk memanfaatkan sumber daya alam. Oleh karena itu, kita memiliki tanggung jawab besar untuk menjaga amanah ini dengan tidak melakukan kerusakan. Menjaga kelestarian lingkungan merupakan salah satu bentuk pengabdian dan rasa syukur kita kepada Allah. Manusia harus patuh dan menghormati petunjuk Allah dalam menjaga kelestarian lingkungan, dan kesadaran ini harus diwujudkan melalui tindakan nyata. Dengan menjaga keseimbangan alam, kita tidak hanya mengikuti perintah-Nya tetapi juga berkontribusi dalam menjaga keberlanjutan dan kesejahteraan bersama. Ini adalah bagian dari tanggung jawab kita sebagai khalifah atau pemimpin di bumi, yang harus dijalankan dengan penuh kesadaran dan kepatuhan terhadap ajaran-Nya.

Untuk mengatasi dampak *Urban Heat Island* diperlukan analisis yang mendalam terhadap fenomena ini. Analisis UHI tidak hanya mencakup pemahaman tentang pola distribusi suhu yang tinggi di perkotaan, tetapi juga melibatkan identifikasi faktor-faktor yang berkontribusi terhadap peningkatan suhu. Faktor-faktor seperti kepadatan penduduk, lahan vegetasi yang berkurang, aktivitas industri, penggunaan lahan, dan karakteristik topografi setempat dapat berpengaruh signifikan terhadap intensitas UHI (Handayanto et al., 2023). Analisis *Urban Heat Island* dapat dilakukan melalui metode analisis data spasial yang memanfaatkan citra satelit. Citra satelit menyediakan data spasial yang luas dan berkelanjutan, memungkinkan para peneliti untuk memantau perubahan suhu permukaan dan distribusi *Urban Heat Island* dalam skala regional atau kota dengan detail yang tinggi. Dengan menggunakan teknik pengolahan citra satelit yang canggih, seperti

analisis citra termal dan spektral, serta algoritma pemrosesan yang tepat, informasi tentang suhu permukaan dan pola *Urban Heat Island* dapat diekstraksi dengan akurat (Larasati et al., 2022).

Dalam penelitian ini, peneliti memilih untuk memanfaatkan *Google Earth Engine* (GEE) sebagai platform untuk analisis citra satelit. GEE dipilih karena menyediakan akses yang mudah dan cepat ke berbagai dataset citra satelit dari sumber-sumber terpercaya, seperti Landsat dan Sentinel-2, yang diperlukan untuk analisis *Urban Heat Island* (Ravanelli et al., 2018). Platform ini juga dilengkapi dengan berbagai alat pemrosesan dan analisis data spasial yang kuat, memungkinkan peneliti untuk melakukan operasi pengolahan citra dan analisis spasial dengan efisien dan efektif. Salah satu keunggulan utama GEE adalah kemampuannya untuk mengintegrasikan data dari berbagai sumber dan melakukan analisis secara skala besar tanpa memerlukan infrastruktur komputasi yang mahal.

Selanjutnya, peneliti bermaksud untuk melakukan penelitian prediksi *Urban Heat Island* menggunakan metode *random forest regression*. Metode ini dipilih karena kemampuannya dalam menangani masalah prediksi dengan dataset yang kompleks dan besar, seperti data citra satelit. Random forest merupakan salah satu teknik yang sangat adaptif dan kuat dalam menghadapi keragaman fitur-fitur yang terdapat dalam dataset, sehingga cocok untuk diterapkan dalam konteks analisis spasial. Sehingga hasil penelitian ini diharapkan mampu memberikan hasil yaitu prediksi UHI yang dapat membantu mengatasi masalah *Urban Heat Island*.

1.2 Rumusan Masalah

Berdasarkan yang telah dijelaskan pada latar belakang, pernyataan masalah yang diambil adalah bagaimana evaluasi performa model prediksi *Urban Heat Island* menggunakan metode *Random Forest Regression* di Kota Malang?

1.3 Batasan Masalah

1. Data yang digunakan adalah data sekunder dari situs *Google Earth Engine* yaitu data satelit Landsat-8, Sentinel-2, dan SRTM.
2. Penelitian ini terbatas pada wilayah di Kota Malang.
3. Evaluasi performa model menggunakan MSE, MAPE, dan R^2 .

1.4 Tujuan Penelitian

Berdasarkan pernyataan masalah yang telah dijelaskan, Tujuan penelitian ini adalah untuk mengukur performa model prediksi metode *Random Forest Regression* pada prediksi *Urban Heat Island* di Kota Malang.

1.5 Manfaat Penelitian

1. Bagi pemerintah Kota Malang, informasi pada penelitian ini berguna untuk digunakan dalam perencanaan dan pembangunan yang akan dilakukan di Kota Malang.
2. Bagi akademisi, penelitian ini akan bermanfaat sebagai kontribusi terhadap literatur ilmiah tentang *Urban Heat Island* dan penerapan metode *Random Forest Regression* dalam pemodelan fenomena lingkungan perkotaan. Hasil penelitian ini dapat menjadi referensi untuk penelitian selanjutnya.

BAB II

KAJIAN PUSTAKA

2.1 Penelitian Terdahulu

Pada penelitian Yao *et al.* (2021) telah dilakukan penelitian prediksi *Urban Heat Island* pada Kota Meihekou di Tiongkok. Penelitian tersebut meneliti tentang prediksi *Surface Urban Heat Island* (SUHI) menggunakan kombinasi dari metode *Random Forest* dengan metode *Monte Carlo* (MC-RFR). Performa metode diuji dengan membandingkan model *Monte Carlo-Random Forest Regression* (MC-RFR), *Random Forest Regression* (RFR), dan *Linear Regression* sebanyak 50 kali menghitung nilai RMSE dan menggunakan R^2 untuk mengevaluasi kemampuan fitting dan akurasi prediksinya. Hasil testnya MC-RFR menghasilkan nilai terbaik dengan nilai tes RMSE 0.822, dan R^2 0.126 kemudian RFR dengan nilai tes RMSE 0.818, dan R^2 0.193.

Penelitian lain tentang *Urban heat Island* oleh Furuya *et al.* (2023) ini bertujuan untuk menerapkan metode *machine learning* prediksi suhu permukaan tanah (LST) berdasarkan variabel sosial ekonomi dan lingkungan, kemudian mengidentifikasi dan mengevaluasi *Surface Urban Heat Island* (SUHI). Data yang digunakan meliputi NDVI, NDBI, dan LST yang kemudian diolah secara statistik untuk menentukan atribut-atributnya. Data LST diperoleh dari citra Landsat 8. Performa metode dievaluasi menggunakan enam algoritma regresi *machine learning* yang berbeda yang dalam empat model yang berbeda. Hasil akhirnya adalah algoritma *Decision Tree* (DT) dalam model pertama (model 1) menunjukkan hasil terbaik dengan akurasi tinggi dalam memprediksi LST.

Penelitian yang dilakukan oleh Dede *et al.* (2019) penelitian ini fokus pada prediksi suhu permukaan menggunakan metode *Artificial Neural Network - Cellular Automata* (ANN-CA). Citra satelit yang dimanfaatkan meliputi Landsat-5 TM, Landsat-8 OLI, dan Sentinel-2 MSI. Proses pengolahan data suhu permukaan melibatkan Radiative Transfer Equation serta konversi nilai brightness temperature menjadi derajat Celsius. Evaluasi kinerja metode ANN-CA dalam memprediksi suhu permukaan menunjukkan nilai *r-square* sebesar 0,80, dengan akurasi keseluruhan 0,63 dan kappa keseluruhan 0,52. Hal ini menghasilkan bahwa model mampu menjelaskan sekitar 80% variabilitas suhu permukaan aktual pada periode 2019 (Dede *et al.*, 2021).

Penelitian oleh Zhao *et al.* (2019), penelitian tersebut mengobservasi tentang data LST dari sensor MODIS di satelit Terra pada hari ke-170 dan ke-181 tahun 2015 di Semenanjung Iberia. Peneliti menggunakan metode *Random Forest Regression* untuk normalisasi temporal. Hasil normalisasi ini menunjukkan peningkatan signifikan dalam akurasi dan bias data LST yang dinormalisasi, sehingga data LST yang dinormalisasi lebih sesuai dengan data LST dari satelit geostasioner MSG.

Tabel 2.1 Penelitian Terkait

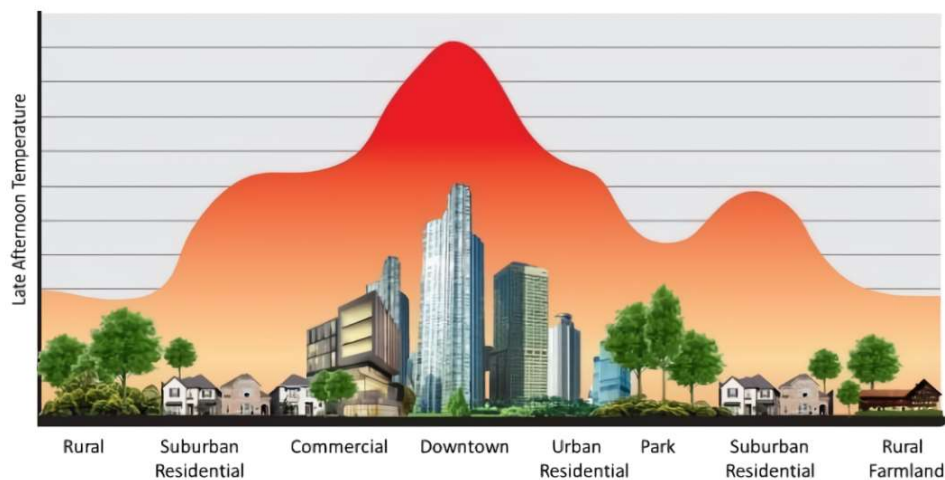
| Peneliti | Judul | Objek dan Metode | Persamaan | Perbedaan | Hasil Penelitian |
|----------------------------|---|---|---|--|--|
| (Yao <i>et al.</i> , 2021) | Predicting Surface Heat Island in Meihokou City, China: A Combination Method of Monte Carlo and Random Forest | Memprediksi SUHI menggunakan MC-RFR menggunakan LST | Penelitian menggunakan data LST dari satelit Landsat dan menggunakan metode RFR | Membandingkan dengan metode MC-RFR dan Linear Regression | Hasil penelitian membandingkan 3 model metode yaitu MC-RFR, RFR, dan Linear Regression. Dengan hasil akhir nilai evaluasi MC-RFR yang paling tinggi. |

| | | | | | |
|-----------------------|--|--|--|--|--|
| (Furuya et al., 2023) | A machine learning approach for mapping surface urban heat island using environmental and socioeconomic variables: a case study in a medium-sized Brazilian city | Memprediksi UHI dengan pendekatan algoritma machine learning menggunakan pendekatan sosial dan ekonomi | Penelitian menggunakan data NDVI, NDBI, dan LST dari satelit Sentinel-2 dan Landsat-8 dan juga menggunakan metode RF | Membandingkan metode KNN, DT, LR, MLP, RF, SVR | Hasil dari penelitian ini adalah performa metode yang diukur dengan menggunakan enam algoritma regresi machine learning yang berbeda dalam empat model yang berbeda. |
| (Dede et al., 2021) | Prediksi Suhu Permukaan Menggunakan Artificial Neural Network-Cellular Automata Di Wilayah Cirebon Dan Sekitarnya | Memprediksi UHI menggunakan metode ANN-Cellular Automata di Cirebon | Penelitian menggunakan data dari satelit Landsat dan Sentinel untuk mendapatkan nilai Surface Temperature (ST) | Menggunakan metode ANN-CA | Menghasilkan Model prediksi suhu permukaan menggunakan metode ANN-CA menunjukkan nilai r-square sebesar 0,80 dengan overall accuracy 0,63 dan overall kappa 0,52. |
| (Zhao et al., 2019) | Normalization of the temporal effect on the MODIS land surface temperature product using random forest regression | Menormalisasikan efek temporal pada data LST dari MODIS menggunakan Random Forest Regression | Penelitian menggunakan metode Random Forest Regression | Menggunakan data satelit MODIS dan Berfokus untuk meneliti normalisasi efek temporal pada data LST | Peningkatan data suhu LST yang dinormalisasikan. Setelah normalisasi terjadi pengurangan perbedaan suhu dan peningkatan konsistensi data LST. |

2.2 Urban Heat Island (UHI)

Urban Heat Island (UHI) merupakan fenomena kenaikan suhu yang signifikan di daerah perkotaan dibandingkan dengan daerah sekitarnya yang lebih pedesaan atau rural. Fenomena ini terjadi karena aktivitas manusia seperti pembangunan gedung, jalan raya, industri, dan transportasi menghasilkan panas yang diserap oleh permukaan kota dan infrastruktur perkotaan (Martin-Vide et al.,

2015). Selain itu, penggunaan material yang menyerap panas seperti beton dan aspal serta minimnya vegetasi juga berkontribusi pada peningkatan suhu di kawasan urban. Akibatnya, suhu udara di perkotaan cenderung lebih tinggi, terutama pada malam hari, karena material tersebut memancarkan panas yang diserap saat siang hari. UHI memiliki dampak yang luas, termasuk peningkatan konsumsi energi untuk pendinginan ruangan, kesehatan masyarakat yang terpapar suhu ekstrem, dan perubahan iklim mikro yang dapat mengganggu ekosistem dan kehidupan fauna serta flora perkotaan (Putra et al., 2021).



Gambar 2.1 Ilustrasi efek dari Urban Heat Island
Sumber: (T & VK, 2018)

Penelitian yang dilakukan (T & VK, 2018) menjelaskan bahwa efek dari *Urban Heat Island* adalah perbedaan suhu yang signifikan antara daerah perkotaan dan daerah pedesaan disekitarnya. Pada gambar 2.1 diilustrasikan bahwa suhu paling tinggi terdapat pada pusat kota (*downtown*) sedangkan semakin ke daerah pinggiran suhu semakin rendah.

Pada penelitian (Mirzaei & Haghghat, 2010) menjelaskan bahwa terjadinya perkembangan pesat dalam pembangunan perkotaan seringkali tidak terencana dengan baik, khususnya dalam hal pengelolaan dampak lingkungan seperti *Urban Heat Island* (UHI). Dalam konteks ini, penelitian tersebut menyoroti adanya tren pembangunan besar-besaran di wilayah perkotaan yang tidak diimbangi dengan strategi mitigasi yang memadai terhadap UHI. Akibatnya, efek samping dari peningkatan suhu yang signifikan di perkotaan, yang disebabkan oleh perubahan tata guna lahan, peningkatan albedo, dan penurunan vegetasi, semakin terasa dan berpotensi memberikan dampak negatif yang serius bagi kesejahteraan penduduk kota serta lingkungan sekitarnya.

Perhitungan *Urban Heat Island* didasarkan pada pengukuran suhu permukaan tanah *Land Surface Temperature* (LST) yang merupakan indikator utama dari fenomena UHI. LST adalah suhu yang diukur langsung dari permukaan tanah atau material yang menutupi permukaan tanah, seperti bangunan, jalan, atau vegetasi. Dalam konteks UHI, LST digunakan sebagai indikator untuk mengukur perbedaan suhu antara wilayah perkotaan dan pedesaan. Perhitungan UHI dinyatakan sebagai berikut (Adi Nugraha & Atmaja, 2020)

$$UHI = T_{Mean} - (\mu + 0.5\alpha) \quad (2.1)$$

Di mana :

T_{Mean} = LST

μ = Nilai rata-rata suhu permukaan

α = Standar deviasi suhu permukaan

2.3 *Land Surface Temperature(LST)*

Land Surface Temperature (LST) atau suhu permukaan tanah adalah suhu yang diukur langsung dari permukaan bumi atau material yang menutupi permukaan tersebut, seperti tanah, vegetasi, air, atau struktur bangunan (Alexander, 2020). LST merupakan parameter penting dalam pemantauan lingkungan dan penelitian iklim karena dapat memberikan informasi tentang karakteristik termal suatu wilayah. Penentuan LST dilakukan dengan menggunakan data citra satelit yang direkam dari sensor termal. Sensor tersebut mengukur radiasi elektromagnetik yang dipancarkan oleh permukaan bumi dalam bentuk panas inframerah.

Untuk memahami di mana peningkatan suhu permukaan terjadi dan mengoptimalkan tata guna lahan serta perencanaan penggunaan lahan, penting untuk memperoleh informasi tentang sebaran Suhu Permukaan Tanah (LST) di suatu daerah (Taripanah & Ranjbar, 2021). Data LST ini dapat memberikan wawasan yang berharga tentang pola suhu yang berbeda-beda di seluruh wilayah, yang merupakan informasi krusial untuk mengambil keputusan dalam perencanaan dan pengelolaan lingkungan.

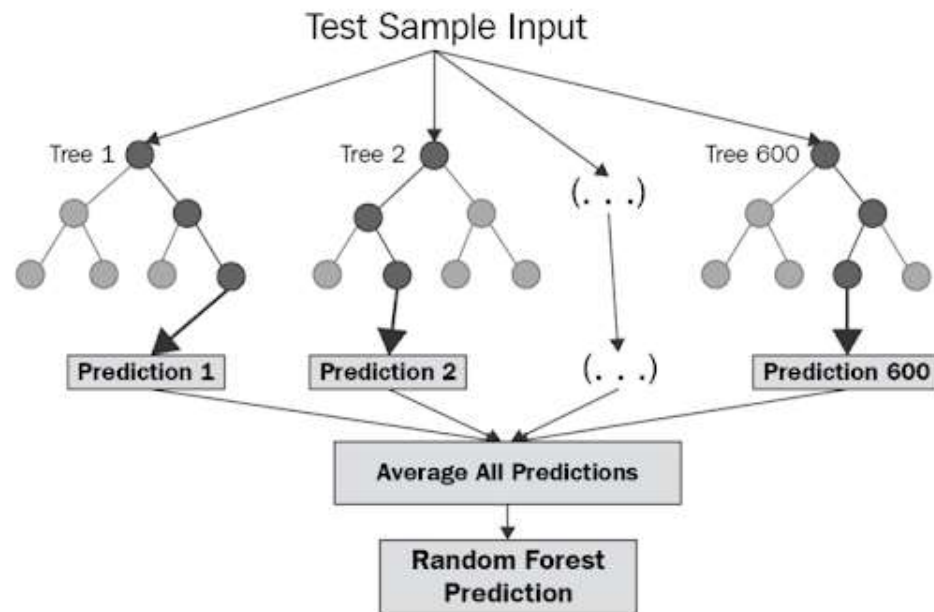
Salah satu sumber data yang dapat menyediakan informasi tentang LST adalah citra satelit. Satelit-satelit seperti Landsat, Sentinel, dan MODIS memiliki sensor termal yang mampu mengukur radiasi inframerah dari permukaan bumi, yang kemudian dapat diolah menjadi data LST. Para peneliti dan pengambil keputusan dapat memperoleh informasi yang diperlukan untuk analisis suhu permukaan dan perencanaan tata guna lahan secara efektif dan efisien.

2.4 *Random Forest*

Random Forest merupakan sebuah metode *ensemble machine learning* yang diperkenalkan pertama kali pada tahun 2001 oleh Leo Breiman (Rose & Hassen, 2019). Metode ini merupakan perkembangan dari *Decision Tree*, yang menggabungkan beberapa pohon prediksi. Setiap pohon dalam kumpulan (*ensemble*) ini bergantung pada nilai-nilai yang acak dan vektor sampel yang independen, serta distribusi yang sama untuk semua pohon dalam kelompok hutan (Schonlau & Zou, 2020). *Decision tree* adalah model prediksi yang menggunakan struktur pohon berhirarki untuk mengambil keputusan berdasarkan serangkaian aturan. Setiap simpul dalam pohon mewakili keputusan berdasarkan fitur-fitur tertentu, dan cabang-cabangnya merepresentasikan kemungkinan hasil keputusan. Proses pembuatan keputusan dimulai dari simpul akar (*root node*) dan berlanjut ke bawah melalui simpul-simpul lainnya hingga mencapai simpul-simpul daun (*leaf node*) yang menunjukkan hasil atau prediksi akhir. *Decision tree* membagi data menjadi subgrup yang lebih kecil berdasarkan fitur-fitur yang relevan, dengan tujuan untuk menghasilkan keputusan yang akurat dan mudah dipahami.

Metode *Random Forest* menggabungkan hasil dari beberapa pohon keputusan, yang masing-masing dibuat dari sampel data yang berbeda dan dengan menggunakan subset acak dari fitur-fitur yang tersedia (Schonlau & Zou, 2020). Setiap pohon dalam *Random Forest* dilatih secara independen, dan hasil prediksi akhir diambil berdasarkan mayoritas suara atau rata-rata dari prediksi yang dihasilkan oleh semua pohon. Pendekatan *ensemble* ini membantu mengatasi kelemahan dari model individual dengan memanfaatkan kekuatan kolektif dari

banyak model yang bekerja bersama-sama. Hal ini membuat Random Forest menjadi metode yang kuat dan serbaguna untuk pemodelan prediksi, yang dapat digunakan dalam berbagai bidang mulai dari klasifikasi hingga regresi.



Gambar 2.2 *Random Forest*
Sumber: builtin.com

Berdasarkan Gambar 2.2 langkah-langkah dalam algoritma *Random Forest* dimulai dengan *input* sampel dari dataset pelatihan secara acak dengan penggantian, yang berarti beberapa sampel dapat dipilih beberapa kali sementara yang lain dapat tidak dipilih sama sekali. Selanjutnya, untuk setiap sampel yang diambil, pohon keputusan dibangun dengan mengikuti prosedur yang sama dengan algoritma *decision tree* biasa. Proses ini melibatkan pemilihan fitur terbaik untuk membagi data pada setiap simpul dan membangun pohon hingga mencapai titik berhenti tertentu, seperti kedalaman maksimum. Setelah sejumlah pohon keputusan dibangun, mereka membentuk ensambel (kumpulan) *tree*. Prediksi dari model

2.5.2 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) adalah matriks evaluasi dengan mengambil nilai kesalahan absolut pada setiap periode, kemudian membaginya dengan nilai observasi aktual pada periode tersebut. Setelahnya, rata-rata dari kesalahan persentase absolut tersebut diambil (Chicco et al., 2021). Nilai MAPE dapat dihitung menggunakan persamaan berikut,

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{X_i - Y_i}{Y_i} \right| \quad (2.3)$$

Di mana :

m = ukuran data

X_i = nilai aktual ke- i

Y_i = nilai prediksi ke- i

Pada perhitungan MAPE, terdapat *range* yang menjadi acuan dalam menilai suatu prediksi, yang dapat dilihat di tabel 2.2 berikut,

Tabel 2.2 Nilai MAPE

| Range MAPE | Arti |
|-------------------|------------------------------------|
| < 10 % | Potensi Model Prediksi Sangat Baik |
| 10 – 20 % | Potensi Model Prediksi Baik |
| 20 – 50 % | Potensi Model Prediksi layak |
| > 50 % | Potensi Model Prediksi Buruk |

2.5.3 R-Squared (R^2)

Koefisien determinasi atau *R-squared* (R^2) adalah metrik evaluasi yang digunakan untuk menunjukkan seberapa baik variabilitas dalam variabel dependen dapat dijelaskan oleh model regresi. Nilai R-squared berkisar antara 0 dan 1, di

mana nilai yang lebih tinggi menunjukkan bahwa model dapat menjelaskan lebih banyak variasi dalam data. Sebuah nilai *R-squared* yang mendekati 1 menunjukkan bahwa model secara baik menjelaskan variasi dalam data, sementara nilai yang mendekati 0 menunjukkan bahwa model tersebut gagal menjelaskan variasi tersebut. Persamaan dari *R-squared* dijelaskan oleh (Gao, 2023) yaitu,

$$R^2 = 1 - \frac{SSR}{SST} \quad (2.4)$$

Di mana :

$$SSR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.5)$$

$$SST = \sum_{i=1}^n (y_i - \bar{y}_i)^2 \quad (2.6)$$

Keterangan :

SSR = jumlah kuadrat residu

SST = jumlah kuadrat total

y_i = nilai aktual ke-i

\hat{y}_i = nilai prediksi ke-i

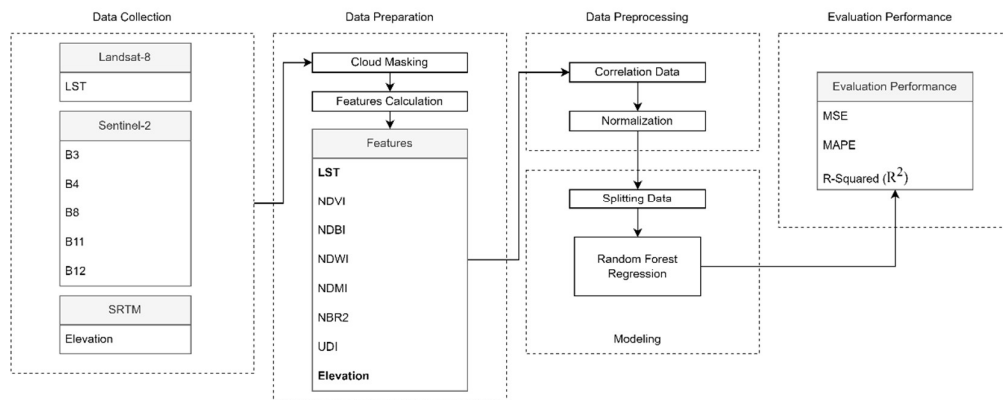
\bar{y}_i = rata-rata nilai aktual

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Desain Sistem

Desain sistem merepresentasikan alur dari penelitian ini secara keseluruhan mulai dari input data hingga menghasilkan prediksi dan hasil performa evaluasi dari model prediksinya. Berikut desain sistem pada gambar 3.1 yang menunjukkan alur pada penelitian ini:



Gambar 3.1 Desain sistem dan penelitian

3.2 Data Collection

Proses pengumpulan data (*Data Collection*) dilakukan untuk mendapatkan data yang digunakan sebagai input data. Pada penelitian ini menggunakan data sekunder yang didapatkan dari situs data citra satelit yaitu *Google Earth Engine*. Data yang diambil pada situs tersebut adalah dari tiga satelit yang digunakan dalam penelitian ini yaitu Sentinel-2, Landsat-8, dan SRTM. Data yang diambil adalah dari tahun 2019 sampai 2023, Rentang waktu ini memungkinkan untuk mendapatkan data dari ketiga satelit yang digunakan dalam penelitian ini, yaitu

Sentinel-2, Landsat-8, dan SRTM. Dengan menggunakan rentang waktu ini, penelitian dapat memperoleh data yang mencakup periode yang cukup panjang untuk analisis yang komprehensif dan memadai terhadap fenomena yang diamati, sekaligus memaksimalkan penggunaan sumber data yang tersedia. Data yang diperoleh dari proses pengumpulan ini sebanyak 4236 data, yang kemudian akan digunakan dalam analisis lebih lanjut untuk memastikan hasil penelitian yang akurat dan informatif.

3.2.1 Satelit Landsat-8

Landsat-8 adalah satelit penginderaan jauh yang diluncurkan oleh *National Aeronautics and Space Administration* (NASA) pada tahun 2013. Landsat-8 ini merupakan satelit paling akhir dari seri landsat dan merupakan yang masih beroperasi dan berfungsi sampai sekarang sebagai keperluan riset untuk kemajuan pengetahuan.

Terdapat beberapa saluran (*band*) pada Landsat-8. Pada penelitian ini, peneliti memilih *band* ST_B10 karena *band* tersebut merekam radiasi termal dari permukaan Bumi, yang digunakan untuk memperkirakan suhu permukaan (*surface temperature*) yang merupakan representasi dari variabel LST (*Land Surface Temperature*).

3.2.2 Satelit Sentinel-2

Sentinel-2 merupakan serangkaian satelit penginderaan jauh yang diluncurkan oleh *European Space Agency* (ESA) sebagai bagian dari program Copernicus. Satelit Sentinel-2 menyediakan data citra multispektral resolusi tinggi

untuk pengamatan Bumi yang luas dan teratur. Diluncurkan pertama kali pada tahun 2015, Sentinel-2 terus beroperasi dan menjadi sumber data yang penting untuk penelitian hingga saat ini.

Dalam penelitian ini, digunakan beberapa saluran (*band*) dari Sentinel-2, yaitu B3, B4, B8, B11, dan B12. Data yang diperoleh dari saluran-saluran ini akan diolah menjadi berbagai fitur dengan menggunakan rumus yang diacu dari referensi. Fitur-fitur yang dihasilkan adalah NDVI, NDBI, NDWI, NDMI, NBR2, dan UDI.

3.2.3 Satelit SRTM

Shuttle Radar Topography Mission (SRTM) adalah satelit penginderaan jauh yang diluncurkan pada tahun 2000 oleh *National Aeronautics and Space Administration* (NASA). Satelit ini bertujuan untuk menghasilkan peta topografi global Bumi dengan resolusi tinggi dan akurasi yang tinggi. SRTM menggunakan teknologi radar interferometri untuk mengukur elevasi permukaan Bumi di seluruh dunia, termasuk wilayah yang sulit dijangkau seperti pegunungan dan daerah terpencil.

Data elevasi dari SRTM tidak masuk langsung ke dalam rumus *Land Surface Temperature* (LST). Namun, data elevasi digunakan sebagai salah satu fitur dalam pemodelan untuk memperkirakan LST. Data elevasi dapat mempengaruhi suhu permukaan melalui efek topografi, di mana daerah dengan elevasi yang lebih tinggi cenderung memiliki suhu yang lebih rendah karena udara yang lebih dingin dan ventilasi udara yang lebih baik. Oleh karena itu, data elevasi dimasukkan ke dalam pemodelan sebagai fitur yang mempengaruhi variabilitas suhu permukaan.

3.3 *Data Preparation*

Data preparation adalah tahap mengolah data mentah yang sudah didapat. Ini melibatkan langkah-langkah seperti pemrosesan *missing values*, pemilihan fitur yang relevan, transformasi variabel, dan pengkodean kategori. Tujuannya adalah untuk menghasilkan dataset yang berkualitas tinggi dan siap digunakan untuk analisis lebih lanjut, seperti pemodelan atau visualisasi. Data preparation membantu meningkatkan akurasi, keandalan, dan interpretabilitas hasil analisis data, serta mengurangi risiko bias atau kesalahan dalam pengambilan keputusan berbasis data.

3.3.1 *Cloud Masking*

Cloud masking adalah proses identifikasi dan penandaan area yang tertutup oleh awan dalam citra satelit atau citra udara. Tujuan utamanya adalah untuk mengidentifikasi area yang tidak dapat digunakan untuk analisis atau pemrosesan lebih lanjut karena cahaya yang diblokir oleh awan. Hasilnya adalah citra yang telah "dibersihkan" dari awan, memungkinkan analisis yang lebih akurat dan relevan.

3.3.2 *Features Calculation*

Pada proses ini yaitu menghitung fitur-fitur yang diperlukan menggunakan rumus matematika yang akan menghasilkan beberapa fitur pada tabel 3.1.

Tabel 3.1 *Dataset Features*

| No | Nama | Keterangan | Satelit |
|----|------|--|------------|
| 1 | LST | Indeks suhu permukaan tanah | Landsat-8 |
| 2 | NDVI | Indeks indentifikasi area hijau | Sentinel-2 |
| 3 | NDBI | Indeks indentifikasi kerapatan distribusi bangunan | Sentinel-2 |

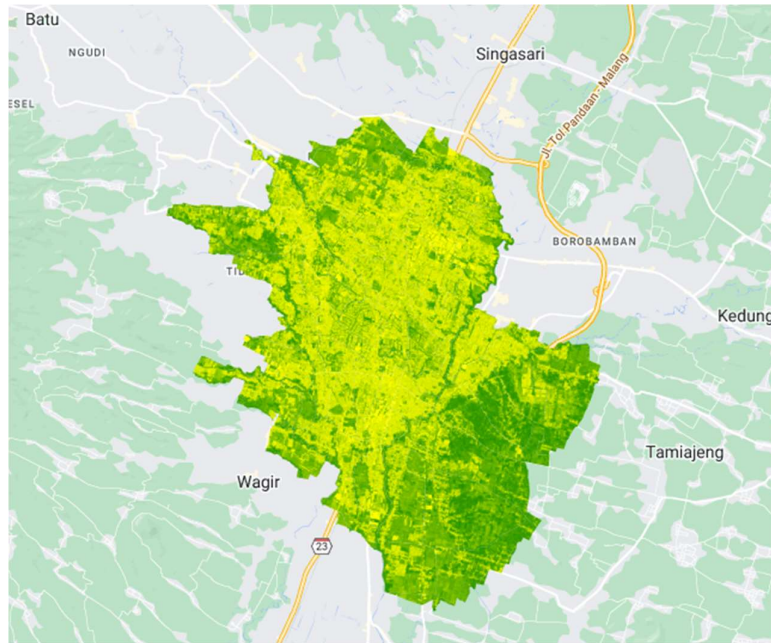
| No | Nama | Keterangan | Satelit |
|----|-----------|--|------------|
| 4 | NDWI | Indeks identifikasi area berair | Sentinel-2 |
| 5 | NDMI | Indeks identifikasi kelembapan tanah | Sentinel-2 |
| 6 | NBR2 | Indeks identifikasi sensitivitas vegetasi terhadap air | Sentinel-2 |
| 7 | UDI | Indeks identifikasi padat area perkotaan | Sentinel-2 |
| 8 | Elevation | Indeks elevasi untuk mengukur ketinggian | SRTM |

3.3.2.1 NDVI

Normalized Difference Vegetation Index (NDVI) merupakan salah satu indeks yang digunakan dalam analisis citra satelit untuk mengukur jumlah dan kondisi vegetasi atau penutupan tanaman di suatu wilayah. NDVI mengukur reflektansi dari permukaan tanah dalam dua rentang panjang gelombang cahaya, yaitu inframerah dekat (NIR) dan merah (RED) dengan rumus sebagai berikut (Rani et al., 2018).

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (3.1)$$

Perbedaan reflektansi antara kedua panjang gelombang ini memberikan informasi tentang keberadaan vegetasi. NDVI dihitung dengan rumus, menghasilkan nilai antara -1 hingga 1. Nilai positif menunjukkan keberadaan vegetasi, di mana nilai yang lebih tinggi menunjukkan vegetasi yang banyak atau lebih padat.



Gambar 3.2 NDVI

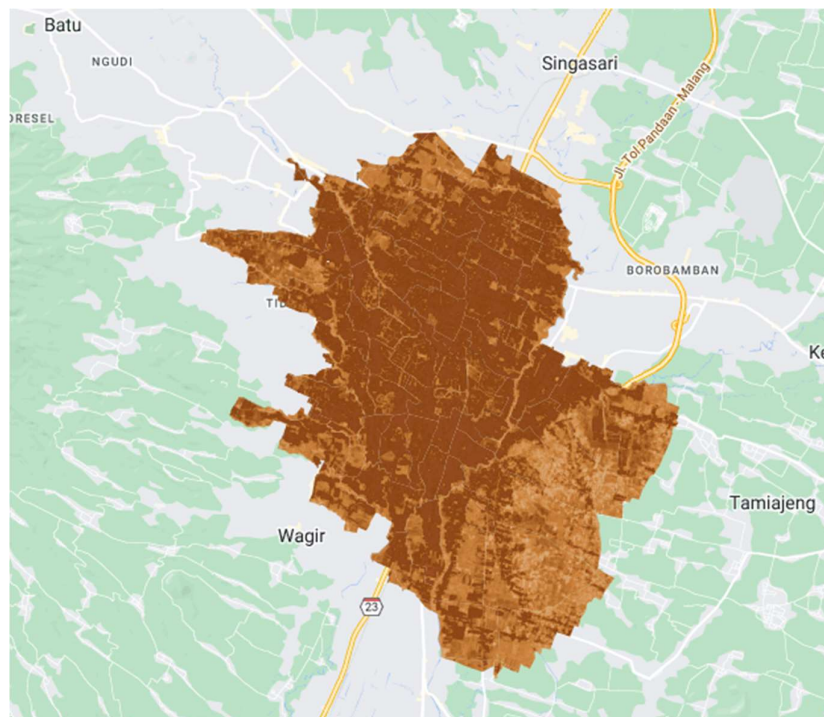
Dalam visualiasi pada gambar 3.2 menunjukkan semakin hijau warnanya maka semakin banyak vegetasinya dan sebaliknya semakin terang maka vegetasi semakin sedikit. Ditunjukkan bahwa daerah pinggi kota lebih hijau dibandingkan dengan daerah tengah kota.

3.3.2.2 NDBI

Normalized Difference Build-up Index (NDBI) merupakan indeks yang digunakan untuk analisis citra satelit untuk mengidentifikasi dan memetakan daerah yang dibangun atau terurbanisasi dalam suatu wilayah. Indeks ini mengukur perbedaan reflektansi antara dua rentang panjang gelombang cahaya, yaitu inframerah dekat (NIR) dan inframerah jauh (SWIR). Rumusnya dinyatakan sebagai berikut (Pranata & Kurniadin, 2021).

$$NDBI = \frac{SWIR2 - NIR}{SWIR2 + NIR} \quad (3.2)$$

Pada persamaan 3.2 diatas SWIR2 berasal dari *band* B12 milik Sentinel-2 yang lebih sensitif terhadap komposisi mineral dan struktur geologis tertentu. NDBI Menghasilkan nilai antara -1 hingga 1. Nilai positif menunjukkan daerah yang dibangun atau terurbanisasi, sementara nilai negatif menunjukkan daerah yang tidak terbangun atau bukan area perkotaan.



Gambar 3.3 NDBI

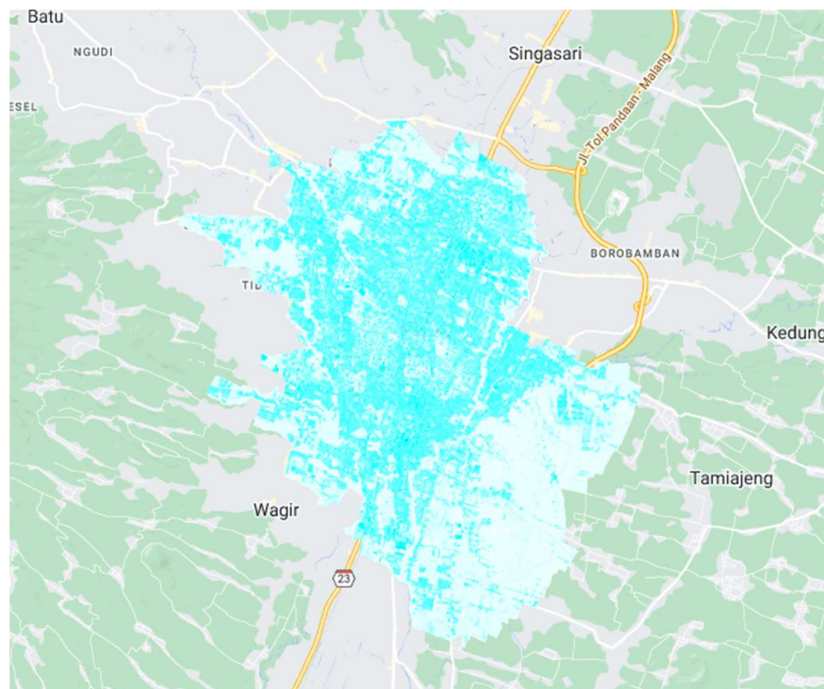
Pada visualiasi pada gambar 3.3 warna coklat menandakan distribusi densitas bangunan yang ada pada wilayah kota. Pada daerah tengah kota warnanya menjadi semakin gelap, sebaliknya daerah pinggir semakin terang karena lebih sedikitnya bangunan.

3.3.2.3 NDWI

Normalized *Difference Water Index* (NDWI) adalah sebuah indeks yang digunakan dalam analisis citra satelit untuk mengidentifikasi dan memetakan daerah yang terdapat air dalam suatu wilayah. Indeks ini berdasarkan pada perbedaan reflektansi antara dua rentang panjang gelombang cahaya, yaitu hijau (GREEN) dan inframerah dekat (NIR). NDWI dihitung dengan persamaan berikut (Permata et al., 2023).

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR} \quad (3.3)$$

Menghasilkan nilai antara -1 hingga 1. Nilai yang tinggi menunjukkan adanya air, sedangkan nilai yang rendah menunjukkan daratan atau area yang tidak memiliki air.



Gambar 3.4 NDWI

Visualisasi yang terlihat pada gambar 3.4 menunjukkan area yang terdapat air. Semakin gelap maka semakin tinggi menunjukkan adanya air pada wilayah tersebut.

3.3.2.4 NDMI

Normalized Difference Moisture Index (NDMI) adalah sebuah indeks yang digunakan dalam analisis citra satelit untuk mengukur tingkat kelembaban atau keberadaan air dalam vegetasi dan tanah. Indeks ini menggabungkan informasi dari dua rentang panjang gelombang cahaya, yaitu inframerah dekat (NIR) dan inframerah tengah (SWIR). Persamaan NDMI dinyatakan sebagai berikut (Malakhov et al., 2020).

$$NDMI = \frac{NIR - SWIR1}{NIR + SWIR1} \quad (3.4)$$

Pada persamaan 3.4 diatas SWIR1 berasal dari band B11 milik Sentinel yang lebih sensitif terhadap kelembaban tanah dan kandungan air dalam vegetasi. Menghasilkan nilai antara -1 hingga 1. Nilai positif menunjukkan keberadaan air atau kelembaban yang tinggi, sedangkan nilai negatif menunjukkan keberadaan air yang rendah atau tanah yang kering.

3.3.2.5 NBR2

Normalized Burn Ratio 2 (NBR2) adalah salah satu indeks yang digunakan untuk menganalisa sensitivitas vegetasi terhadap air. Indeks ini menggabungkan informasi dari dua *band* berbeda, yaitu inframerah jauh (SWIR) dan SWIR2. SWIR1 dan SWIR2 memiliki panjang gelombang yang berbeda, yang masing-

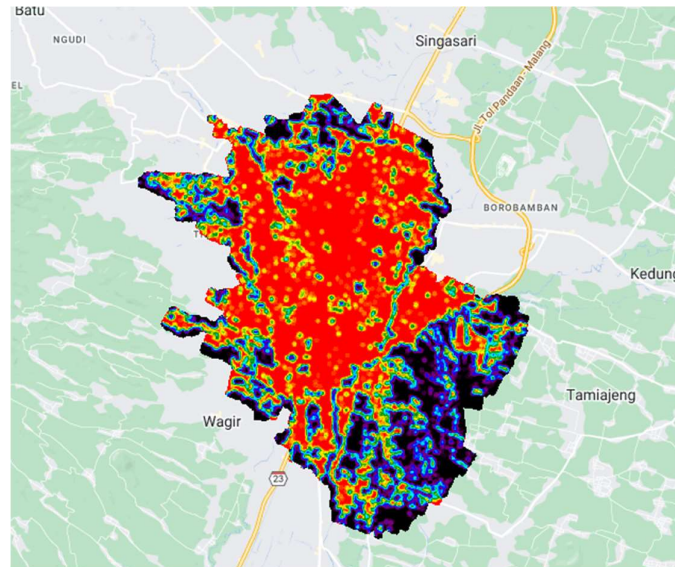
masing dapat memberikan informasi tambahan tentang karakteristik lahan dan vegetasi. NBR2 dihitung dengan rumus berikut (Dvorakova et al., 2021).

$$NBR2 = \frac{SWIR1 - SWIR2}{SWIR1 + SWIR2} \quad (3.5)$$

Pada persamaan 3.5 dapat dilihat menggunakan inframerah jauh yang berbeda. SWIR1 cenderung lebih sensitif terhadap kepadatan vegetasi dan kondisi lahan, sementara SWIR2 memiliki sensitivitas yang lebih tinggi terhadap perubahan dalam kandungan air dan tekstur tanah. Oleh karena itu, NBR2 dapat memberikan informasi yang lebih komprehensif tentang kerusakan lahan atau vegetasi akibat kebakaran, karena memanfaatkan karakteristik kedua saluran SWIR tersebut.

UDI

Urban Density Index (UDI) merupakan metrik yang digunakan dalam perencanaan dan desain perkotaan untuk mengevaluasi populasi yang tinggal di suatu wilayah perkotaan. Indeks ini menggambarkan seberapa padat populasi dan infrastruktur perkotaan dalam suatu area tertentu. Pada gambar 3.5 memvisualisasikan intensitas UDI. Warna cerah merah menunjukkan kepadatan yang ada pada wilayah tersebut.



Gambar 3.5 UDI

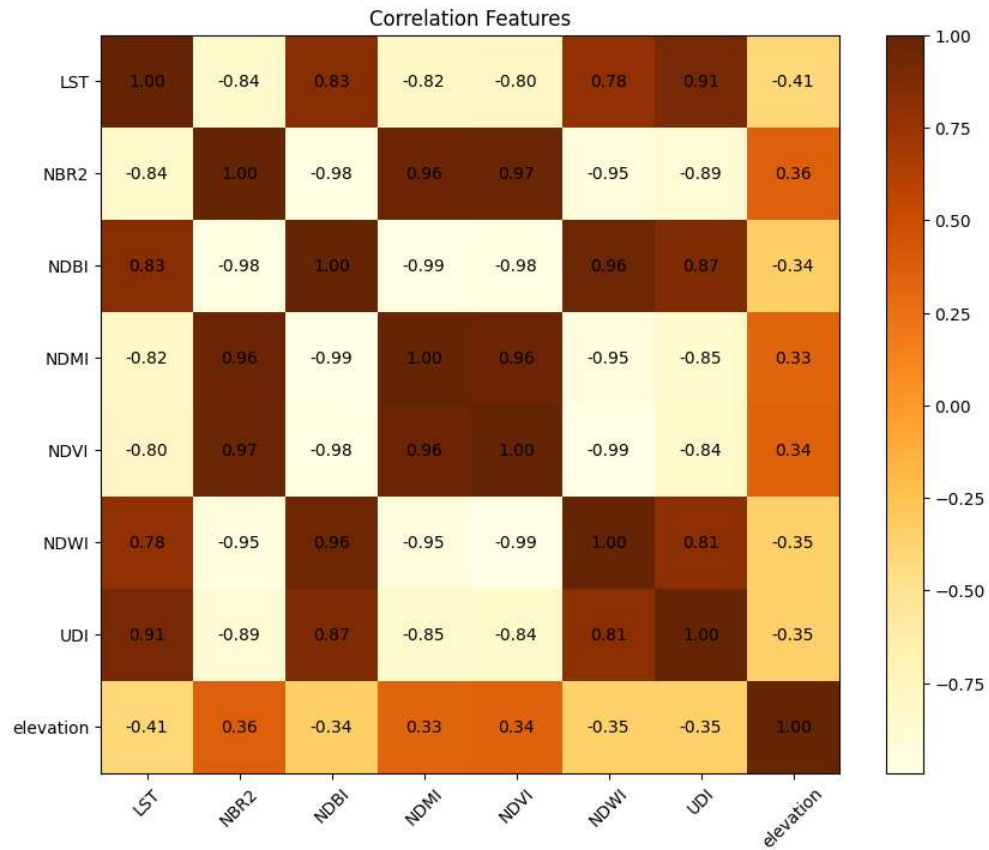
UDI sering digunakan bersama dengan metrik lain, seperti *Green Density Index* (GDI), untuk memberikan gambaran yang lebih holistik tentang lingkungan perkotaan. *Green Density Index*, di sisi lain, adalah ukuran yang menilai jumlah area hijau dalam wilayah perkotaan, yang dihitung dengan menggunakan indeks spektral seperti NDWI, NDBI, dan NBR2. Dari nilai piksel hasil perhitungan tersebut, kemudian diambil rata-rata untuk mendapatkan nilai *Green Density Index*.

3.4 *Data Preprocessing*

Dataset yang telah dimuat kemudian dianalisis kembali untuk mempermudah perhitungan lebih lanjut. Langkah awal dalam pengolahan data adalah menginisialisasi setiap variabel yang akan dihitung. *Preprocessing data* adalah serangkaian proses yang dilakukan untuk mengumpulkan, membersihkan, mengatur, menganalisis, dan menafsirkan data dengan tujuan agar data dapat dilatih dan diuji pada tahap berikutnya.

3.4.1 *Data Correlation*

Korelasi data mengacu pada hubungan statistik antara dua atau lebih variabel, yang menunjukkan sejauh mana variabel-variabel tersebut bergerak bersamaan satu sama lain. Dalam analisis korelasi, penting untuk memahami apakah terdapat hubungan positif, negatif, atau tidak ada hubungan antara variabel-variabel tersebut. Korelasi positif terjadi ketika nilai variabel satu meningkat bersamaan dengan nilai variabel lainnya, sementara korelasi negatif terjadi ketika nilai satu variabel meningkat sementara nilai variabel lainnya menurun. Sebaliknya, korelasi yang lemah atau tidak ada korelasi menunjukkan bahwa tidak ada hubungan linier yang jelas antara variabel-variabel tersebut. Analisis korelasi sangat penting dalam pengambilan keputusan karena membantu mengidentifikasi pola-pola dan tren dalam data, serta memungkinkan kita untuk memprediksi atau menjelaskan perilaku variabel-variabel tersebut berdasarkan hubungan yang diamati.



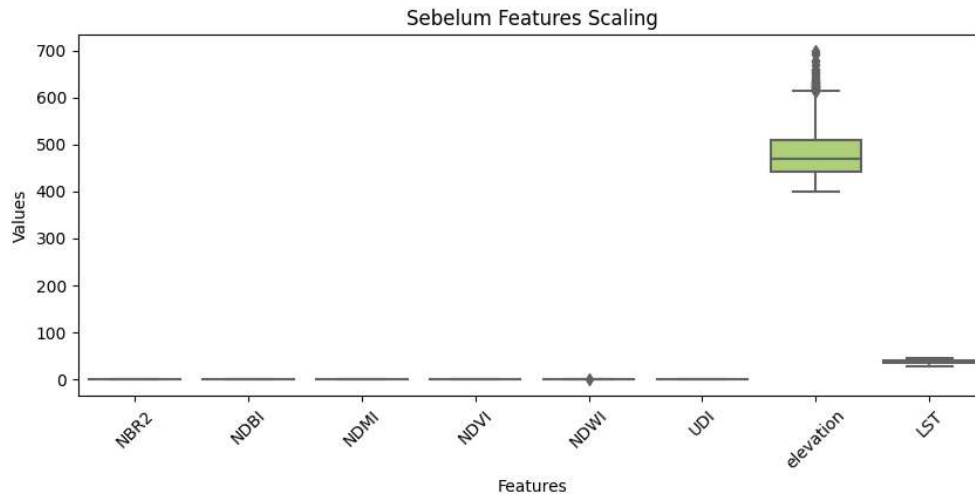
Gambar 3.6 Data Correlation

Gambar 3.6 menunjukkan korelasi antar data dalam bentuk heatmap. Indikator warna pada gambar tersebut menunjukkan warna gelap dengan nilai 1 jika positif dan warna cerah menunjukkan korelasi nilai -1 yang berarti negatif. Jika nilai pada korelasi mendekati 0 maka fitur tersebut tidak memiliki hubungan korelasi yang kuat.

3.4.2 Normalisasi

Normalisasi data merupakan proses mengubah skala nilai-nilai variabel dalam dataset sehingga mereka berada dalam rentang yang seragam. Tujuan

normalisasi adalah untuk menghilangkan perbedaan skala antar variabel, sehingga setiap variabel memiliki pengaruh yang seimbang terhadap hasil analisis.



Gambar 3.7 Data sebelum normalisasi

Seperti yang ditampilkan pada gambar 3.7 dapat terlihat bahwa dataset tidak berada dalam rentang yang seragam. Nilai dari *elevation* berada jauh lebih tinggi dengan yang lainnya, oleh karena itu diperlukan proses normalisasi agar datanya menjadi seragam.

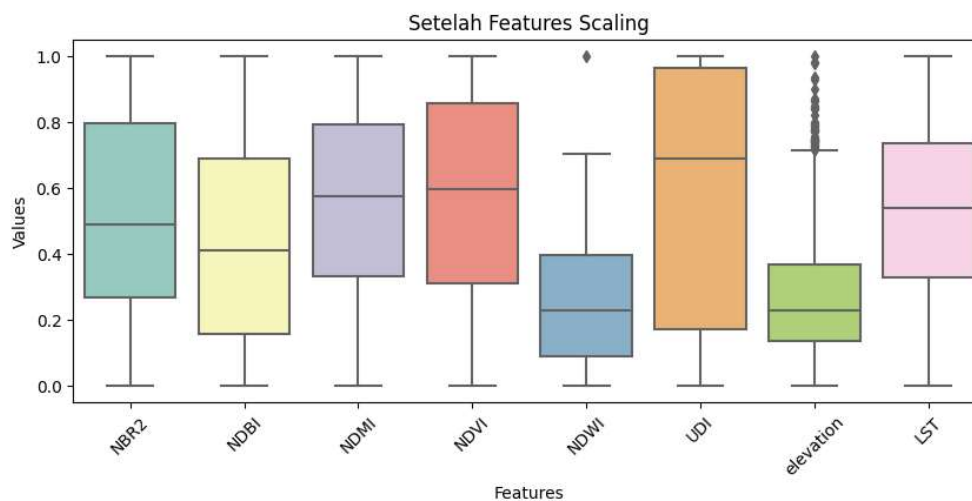
Normalisasi dengan *MinMaxScaler* adalah salah satu teknik normalisasi yang umum digunakan dalam analisis data. Metode ini bertujuan untuk mengubah distribusi nilai variabel sehingga memiliki *mean* (rerata) 0 dan standar deviasi 1. Proses ini dilakukan dengan mengurangi nilai setiap observasi dengan rerata dari seluruh sampel, kemudian dibagi dengan deviasi standar dari seluruh sampel. Berikut adalah persamaan *MinMaxScaler*:

$$x_{min-max} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3.6)$$

Dengan demikian, setelah normalisasi, variabel memiliki skala yang seragam dan terpusat di sekitar 0 hingga 1. Berikut pada tabel 3.1 adalah contoh perbandingan dari data sebelum dan sesudah dilakukan normalisasi menggunakan persamaan *MinMaxScaler*.

Tabel 3.2 Data sampel contoh hasil normalisasi

| Data Sebelum Normalisasi | | | | | | | |
|--------------------------|----------|----------|----------|----------|----------|-------|-----------|
| Index | NBR2 | NDBI | NDMI | NDVI | NDWI | UDI | Elevation |
| 0 | -0.042 | 0.199 | -0.158 | -0.033 | 0.009 | 0.931 | 491 |
| 1 | -0.065 | 0.283 | -0.222 | -0.017 | 0.008 | 0.966 | 466 |
| 2 | -0.018 | 0.173 | -0.155 | -0.001 | 0.001 | 0.931 | 424 |
| 3 | 0.028 | -0.009 | -0.019 | -0.029 | 0.032 | 0.966 | 431 |
| 4 | -0.018 | 0.237 | -0.22 | -0.02 | 0.012 | 0.966 | 463 |
| Data Sesudah Normalisasi | | | | | | | |
| Index | NBR2 | NDBI | NDMI | NDVI | NDWI | UDI | Elevation |
| 0 | 0.105788 | 0.787196 | 0.290155 | 0.071800 | 0.593023 | 0.931 | 0.302013 |
| 1 | 0.059880 | 0.862940 | 0.207254 | 0.088450 | 0.592248 | 0.966 | 0.218121 |
| 2 | 0.153693 | 0.763751 | 0.294041 | 0.105099 | 0.586822 | 0.931 | 0.077181 |
| 3 | 0.245509 | 0.599639 | 0.470207 | 0.075963 | 0.610853 | 0.966 | 0.100671 |
| 4 | 0.153693 | 0.821461 | 0.209845 | 0.085328 | 0.595349 | 0.966 | 0.208054 |



Gambar 3.8 Data setelah normalisasi

Hasil dari normalisasi menjadikan rentang nilainya jadi berdekatan, seperti pada contoh nilai Elevation yang awalnya sangat tinggi, sekarang berada dalam rentang yang sama dengan fitur lainnya. Untuk visualisasi grafik rentang data hasil normalisasi dapat dilihat pada gambar 3.8 menunjukkan bahwa seluruh data berada pada rentang 0 sampai 1.

3.5 Modeling

Pada tahap ini merupakan tahap pembuatan model prediksi. Pertama data akan dipisah (*splitting*) menjadi model latihan (*training*) dan untuk testing. Kemudian kedua data tersebut akan diprediksi seperti yang mana langkahnya sudah dijelaskan pada gambar 3.1.

3.5.1 Splitting Data

Splitting data merupakan tahap yang dilakukan setelah tahap *Preprocessing* dilakukan, Tahap ini adalah melakukan pembagian data menjadi dua yaitu data *training* dan data *testing* yang dimana pada konteks penelitian ini *splitting data* dibagi menjadi tiga rasio perbandingan yaitu 70% data latih, 80% data latih, dan 90% data latih. Rasio 70% menghasilkan 3377 data *train* dan 1448 data *test*. Rasio 80% menghasilkan 3860 data *train* dan 965 data *test*. Rasio 90% menghasilkan 4342 data *train* dan 483 data *test*.

3.5.2 Training Model

Setelah proses *splitting data* selesai, tahap berikutnya adalah melatih model menggunakan metode *Random Forest Regression*. Tahap ini disebut sebagai tahap pelatihan model atau *training model*. Pada tahap ini, model akan belajar dari data

latih yang telah disiapkan sebelumnya. Proses pelatihan dilakukan dengan memberikan data *input* (fitur) dan *output* yang sesuai kepada algoritma *Random Forest Regression*.

Classification and Regression Trees (CART) merupakan istilah yang digunakan dalam pembentukan pohon keputusan dalam algoritma *Random Forest*. CART menggunakan aturan partisi rekursif untuk membagi data menjadi subset yang semakin kecil dan homogen berdasarkan nilai-nilai fitur. Proses ini dilakukan secara berulang untuk membentuk struktur pohon yang kompleks. Pada setiap node dalam pohon keputusan, CART memilih fitur dan nilai pemisah yang paling baik untuk memaksimalkan homogenitas subset yang dihasilkan. Hasil dari CART adalah pohon klasifikasi jika variabelnya adalah kategorik dan akan menghasilkan pohon regresi jika variabelnya berupa data kontinu seperti pada konteks penelitian ini.

3.5.2.1 *Bootstrap sampling*

Dalam *bootstrap sampling*, setiap pohon dibangun menggunakan subset acak dari data latih. Hal ini berarti bahwa setiap sampel data memiliki peluang untuk dipilih beberapa kali atau tidak sama sekali dalam subset untuk setiap pohon. Teknik ini membantu dalam mengurangi overfitting karena setiap pohon dibangun dengan data yang sedikit berbeda. Berikut contoh *Bootstrap sampling* menggunakan 6 baris sampel data,

Tabel 3.3 Data sampel untuk *bootstrap sampling*

| IDX | NBR2 | NDBI | NDMI | NDVI | NDWI | UDI | Elevation | LST |
|-----|----------|----------|----------|----------|----------|-------|-----------|--------|
| 0 | 0.105788 | 0.787196 | 0.290155 | 0.071800 | 0.593023 | 0.931 | 0.302013 | 40.163 |
| 1 | 0.059880 | 0.862940 | 0.207254 | 0.088450 | 0.592248 | 0.966 | 0.218121 | 39.746 |
| 2 | 0.153693 | 0.763751 | 0.294041 | 0.105099 | 0.586822 | 0.931 | 0.077181 | 40.655 |

| IDX | NBR2 | NDBI | NDMI | NDVI | NDWI | UDI | Elevation | LST |
|-----|----------|----------|----------|----------|----------|-------|-----------|--------|
| 3 | 0.245509 | 0.599639 | 0.470207 | 0.075963 | 0.610853 | 0.966 | 0.100671 | 42.532 |
| 4 | 0.153693 | 0.821461 | 0.209845 | 0.085328 | 0.595349 | 0.966 | 0.208054 | 43.758 |
| 5 | 0.061876 | 0.912534 | 0.132124 | 0.104058 | 0.570543 | 0.966 | 0.104027 | 37.662 |

Tabel 3.4 Contoh hasil *bootstrap sampling*

| IDX | NBR2 | NDBI | NDMI | NDVI | NDWI | UDI | Elevation | LST |
|-----|----------|----------|----------|----------|----------|-------|-----------|--------|
| 1 | 0.059880 | 0.862940 | 0.207254 | 0.088450 | 0.592248 | 0.966 | 0.218121 | 39.746 |
| 1 | 0.059880 | 0.862940 | 0.207254 | 0.088450 | 0.592248 | 0.966 | 0.218121 | 39.746 |
| 0 | 0.105788 | 0.787196 | 0.290155 | 0.071800 | 0.593023 | 0.931 | 0.302013 | 40.163 |
| 3 | 0.245509 | 0.599639 | 0.470207 | 0.075963 | 0.610853 | 0.966 | 0.100671 | 42.532 |
| 4 | 0.153693 | 0.821461 | 0.209845 | 0.085328 | 0.595349 | 0.966 | 0.208054 | 43.758 |
| 5 | 0.061876 | 0.912534 | 0.132124 | 0.104058 | 0.570543 | 0.966 | 0.104027 | 37.662 |

Tabel 3.4 merupakan hasil *Bootstrap sampling*. Dalam tabel tersebut, kita dapat melihat beberapa baris yang muncul lebih dari sekali. Dapat dilihat bahwa *index 1* muncul dua kali dan *index 2* tidak muncul pada *Bootstrap* tersebut. *Bootstrap sampling* menginkan terjadi pengulangan sampel, karena setiap kali baris data diambil, baris data tersebut dikembalikan sehingga bisa diambil secara acak kembali pada pengulangan berikutnya. Proses *Bootstrap sampling* dilakukan berdasarkan seberapa banyak *parameter tree* yang diatur pada saat proses membuat model.

3.5.2.2 Membuat *Decision Tree*

Pembuatan pohon keputusan(*Decision Tree*) dibuat dari hasil proses *Bootstrap sampling* sebelumnya. Proses pembuatan *Decision Tree* ini dimulai dengan melakukan *split* dari data *Bootstrap* yang telah dibuat. Proses *split* dilakukan secara berulang sampai mencapai ukuran minimum yang telah ditentukan.

CART (*Classification and Regression Trees*) memiliki metode untuk menghitung kualitas *split* atau pembagian data dalam kedua konteksnya: klasifikasi dan regresi. Dalam klasifikasi, CART menggunakan *Gini impurity* atau *Information Gain* untuk menentukan kualitas split. Sedangkan dalam konteks regresi, CART menggunakan *Variance Reduction* untuk mengukur kualitas *split*. *Variance Reduction* dapat dihitung menggunakan *Mean Squared Error* (MSE).

Berikut merupakan contoh perhitungan *split decision* menggunakan *Variance Reduction* dengan *Mean Squared Error*(MSE) berdasarkan data hasil *Bootstrap* tabel 3.4,

Pada awal pembuatan *Decision Tree*, kita menentukan *threshold* atau batas pada splitnya. *Threshold* dapat ditentukan dengan berbagai cara seperti *mean*, *median*, dan lainnya. Misal pada contoh disini kita menggunakan fitur NDBI dan misal nilainya adalah 0.8043285.

Maka data akan terbagi menjadi dua subset:

Subset 1 (NDBI \leq 0.8043285):

Tabel 3.5 Subset 1 node kiri

| IDX | NDBI | LST |
|-----|----------|--------|
| 0 | 0.787196 | 40.163 |
| 3 | 0.599639 | 42.532 |

Subset 2 (NDBI $>$ 0.8043285):

Tabel 3.6 Subset 2 node kanan

| IDX | NDBI | LST |
|-----|----------|--------|
| 1 | 0.862940 | 39.746 |
| 1 | 0.862940 | 39.746 |
| 4 | 0.821461 | 43.758 |
| 5 | 0.912534 | 37.662 |

Pada masing masing node hitung nilai MSE-nya:

$$MSE_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{Y}_i)^2 \quad (3.7)$$

Di mana :

MSE_n = Nilai MSE pada *tree* ke-n.

N = Jumlah sampel pada *tree* ke-n.

X_i = Nilai sampel ke-I pada pohon ke-n.

\bar{Y}_i = = Nilai rata-rata sampel pohon ke-n.

Maka perhitungannya sebagai berikut:

$$\bar{Y}(1) = \frac{40.163 + 42.532}{2} = 41.3475 \quad (3.8)$$

$$MSE(1) = \frac{1}{2} ((40.163 - 41.3475)^2 + (42.532 - 41.3475)^2) = \mathbf{1.401} \quad (3.9)$$

$$\bar{Y}(2) = \frac{439.746 + 39.746 + 43.758 + 37.662}{4} = 40.228 \quad (3.10)$$

$$MSE(2) = \frac{1}{4} ((39.746 - 40.228)^2 + (39.746 - 40.228)^2 + (43.758 - 40.228)^2 + (37.662 - 40.228)^2) = \mathbf{4.889} \quad (3.11)$$

Seperti yang diketahui bahwa nilai MSE yang paling kecil adalah nilai yang paling baik. Maka untuk selanjutnya subset atau node yang terpilih adalah MSE(1) kemudian langkah-langkah yang sudah dikerjakan akan diulangi kembali sampai sampelnya tersisa satu. Tahap pembuatan *Decision Tree* ini dilakukan pada semua *Bootstrap samples* yang telah dibuat, sehingga akan banyak *tree* yang tercipta yang dari banyaknya itu kita dapat menyebutnya sebagai hutan(*forest*).

3.5.2.3 Agregasi Hasil

Tahap agregasi hasil adalah langkah akhir dalam algoritma *Random Forest*, di mana prediksi dari semua *tree* yang telah dibangun digabungkan untuk menghasilkan prediksi akhir. Proses ini berbeda untuk tugas regresi dan klasifikasi. Setiap pohon dalam hutan memberikan prediksi individual untuk setiap sampel dalam data uji atau data baru. Misalnya, jika ada 100 *tree* setiap sampel akan memiliki 100 prediksi berbeda (satu dari masing-masing *tree*). Dalam konteks regresi, hasil akhir agregasi dilakukan dengan menghitung rata-rata dari semua prediksi individual tersebut (Biau & Scornet, 2016). Selain agregasi rata-rata (*mean*), kita dapat menggunakan berbagai cara lain salah satunya adalah menggunakan median, hasil bisa saja lebih baik untuk meningkatkan kekokohan (*robustness*) algoritimanya, hasilnya bisa juga lebih buruk tergantung pada skenario dan skemanya (Roy & Larocque, 2012).

Misalnya, jika ada lima *tree* memberikan prediksi 39.5, 40.2, 39.8, 40.0, dan 39.7, prediksi akhirnya adalah rata-rata dari nilai-nilai tersebut:

$$\text{Prediksi mean} = \frac{39.5 + 40.2 + 39.8 + 40.0 + 39.7}{5} = 39.84 \quad (3.12)$$

Kemudian untuk contoh agregasi dengan median, adalah mengambil nilai tengah dari data yang diurutkan. Maka jika diurutkan: 39.5, 39.7, 39.8, 40.0, 40.2. Sehingga ditemukan nilai tengahnya adalah 39.8.

3.6 Evaluation Performance

Tahapan terakhir pada penelitian ini adalah melakukan *evaluation performance*. Evaluasi merupakan langkah terakhir dan sangat penting untuk menilai seberapa baik performa dari model prediksi yang telah dibuat. Evaluasi performa pada penelitian ini menggunakan MSE, MAPE, dan R^2 .

Berikut contoh perhitungan MSE,

$$MSE = \frac{1}{4} \sum (3.0 - 2.5)^2 + (4.5 - 4.0)^2 + (2.0 - 2.5)^2 + (5.0 - 4.5)^2 = 0.25 \quad (3.13)$$

Berikut contoh perhitungan MAPE,

$$MAPE = \frac{1}{4} \sum_i \left(\left| \frac{100 - 90}{100} \right| + \left| \frac{150 - 160}{150} \right| + \left| \frac{200 - 210}{200} \right| + \left| \frac{250 - 240}{250} \right| \right) \times 100\% \quad (3.14)$$

$$= 6.42\%$$

3.7 Skenario

Dalam penelitian ini, terdapat serangkaian skenario uji yang telah dirancang untuk mengevaluasi kinerja model. Skenario uji ini dirancang dengan mempertimbangkan variasi rasio dari data latih dan testing serta *tuning parameter* yang telah ditentukan. Berikut skenario uji pada penelitian ini,

Tabel 3.7 Skenario uji

| Skenario | Agregasi Hasil | Ratio data training : data testing | Tuning Parameter |
|----------|-----------------|---------------------------------------|--|
| 1 | Agregasi Mean | 70:30 | <i>Tree states :</i> 50, 100, 250, 500 <i>Max depth tree :</i> 5, 10, 20 <i>Bootstrap Samples :</i> 25%, 50%, 70% |
| | | 80:20 | |
| | | 90:10 | |
| 2 | Agregasi Median | 70:30 | |
| | | 80:20 | |
| | | 90:10 | |

Tabel 3.7 menunjukkan bahwa terdapat dua skenario agregasi akhir, lalu masing-masing terdapat rasio dari *splitting data* untuk *data training: data testing* ada tiga. Kemudian pada *tuning parameter* ada *tree states* sebagai parameter yang mengatur banyaknya *tree*, *Max depth* untuk mengatur kedalaman *tree*, dan *bootstrap samples* untuk mengatur persentase banyaknya subset acak yang diambil pada proses *bootstrap samples*. Dari agregasi, rasio dan *tuning parameter* tersebut maka total yang dihasilkan diakhir akan menjadi 216 varian model. Dengan menggunakan kombinasi pada skenario ini bertujuan untuk mengamati dan memahami bagaimana variasi tersebut dalam model *Random Forest Regression* mempengaruhi kinerja dan akurasi prediksi sehingga akan terlihat variasi model terbaik di hasil akhirnya.

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Hasil Uji Coba

Pada bagian ini, akan membahas hasil evaluasi performa model dengan menggunakan metode Random Forest Regression berdasarkan variasi penggunaan agregasi, rasio dan tuning parameter sesuai dengan skenario pengujian yang telah dijabarkan pada tabel 3.7. Setiap skenario diuji untuk mengukur seberapa efektif model dapat memprediksi hasil dengan parameter yang berbeda-beda.

4.1.1 Hasil Uji Coba Skenario 1

Uji coba pada skenario 1 ini *Random Forest Regression* diuji dengan agregasi akhir *mean* dengan sub-skenario yaitu pembagian rasio data training dan testing 70:30, 80:20, dan 90:10 yang masing-masing diuji dengan tuning parameter yang sudah diatur. Berikut merupakan hasil uji skenario 1 secara keseluruhan,

Tabel 4.1 Hasil skenario 1 agregasi mean

| Model | Ratio train | Trees | Max Depth | Bootstrap Samples | MSE | MAPE | R2 |
|-------|-------------|-------|-----------|-------------------|----------|----------|----------|
| 1 | 70% | 50 | 5 | 0.25 | 1.892515 | 2.949395 | 0.839317 |
| 2 | | 50 | 5 | 0.5 | 1.918871 | 2.964438 | 0.837079 |
| 3 | | 50 | 5 | 0.7 | 1.924423 | 2.965619 | 0.836608 |
| 4 | | 50 | 10 | 0.25 | 1.790709 | 2.869506 | 0.847961 |
| 5 | | 50 | 10 | 0.5 | 1.800773 | 2.860813 | 0.847106 |
| 6 | | 50 | 10 | 0.7 | 1.845312 | 2.873408 | 0.843325 |
| 7 | | 50 | 20 | 0.25 | 1.810982 | 2.864149 | 0.846239 |
| 8 | | 50 | 20 | 0.5 | 1.816525 | 2.854883 | 0.845769 |
| 9 | | 50 | 20 | 0.7 | 1.883122 | 2.88078 | 0.840114 |
| 10 | | 100 | 5 | 0.25 | 1.889753 | 2.95041 | 0.839551 |
| 11 | | 100 | 5 | 0.5 | 1.916351 | 2.959067 | 0.837293 |
| 12 | | 100 | 5 | 0.7 | 1.911508 | 2.953256 | 0.837704 |
| 13 | | 100 | 10 | 0.25 | 1.792434 | 2.866813 | 0.847814 |

| Model | Ratio train | Trees | Max Depth | Bootstrap Samples | MSE | MAPE | R2 | |
|--------------|------------------------|--------------|----------------------|------------------------------|------------|-------------|-----------|----------|
| 14 | | 100 | 10 | 0.5 | 1.793384 | 2.851547 | 0.847734 | |
| 15 | | 100 | 10 | 0.7 | 1.812734 | 2.851466 | 0.846091 | |
| 16 | | 100 | 20 | 0.25 | 1.799757 | 2.859444 | 0.847192 | |
| 17 | | 100 | 20 | 0.5 | 1.796866 | 2.84189 | 0.847438 | |
| 18 | | 100 | 20 | 0.7 | 1.844405 | 2.853122 | 0.843402 | |
| 19 | | 250 | 5 | 0.25 | 1.900374 | 2.951531 | 0.83865 | |
| 20 | | 250 | 5 | 0.5 | 1.902958 | 2.947175 | 0.83843 | |
| 21 | | 250 | 5 | 0.7 | 1.907105 | 2.946122 | 0.838078 | |
| 22 | | 250 | 10 | 0.25 | 1.786986 | 2.853201 | 0.848277 | |
| 23 | | 250 | 10 | 0.5 | 1.782641 | 2.849553 | 0.848646 | |
| 24 | | 250 | 10 | 0.7 | 1.7915 | 2.847967 | 0.847894 | |
| 25 | | 250 | 20 | 0.25 | 1.796028 | 2.849435 | 0.847509 | |
| 26 | | 250 | 20 | 0.5 | 1.78716 | 2.836089 | 0.848262 | |
| 27 | | 250 | 20 | 0.7 | 1.811874 | 2.844523 | 0.846164 | |
| 28 | | 500 | 5 | 0.25 | 1.900752 | 2.95245 | 0.838618 | |
| 29 | | 500 | 5 | 0.5 | 1.898541 | 2.94398 | 0.838805 | |
| 30 | | 500 | 5 | 0.7 | 1.908213 | 2.944972 | 0.837984 | |
| 31 | | 500 | 10 | 0.25 | 1.784713 | 2.853139 | 0.84847 | |
| 32 | | 500 | 10 | 0.5 | 1.773741 | 2.843188 | 0.849401 | |
| 33 | | 500 | 10 | 0.7 | 1.791583 | 2.844961 | 0.847886 | |
| 34 | | 500 | 20 | 0.25 | 1.789078 | 2.846774 | 0.848099 | |
| 35 | | 500 | 20 | 0.5 | 1.777786 | 2.83372 | 0.849058 | |
| 36 | | 500 | 20 | 0.7 | 1.807217 | 2.84179 | 0.846559 | |
| 37 | | 80% | 50 | 5 | 0.25 | 1.958211 | 2.954266 | 0.829548 |
| 38 | | | 50 | 5 | 0.5 | 1.942414 | 2.972414 | 0.830923 |
| 39 | | | 50 | 5 | 0.7 | 1.963665 | 2.966129 | 0.829074 |
| 40 | | | 50 | 10 | 0.25 | 1.851713 | 2.862913 | 0.838819 |
| 41 | | | 50 | 10 | 0.5 | 1.826899 | 2.854627 | 0.840978 |
| 42 | | | 50 | 10 | 0.7 | 1.832136 | 2.884726 | 0.840523 |
| 43 | | | 50 | 20 | 0.25 | 1.858326 | 2.858309 | 0.838243 |
| 44 | | | 50 | 20 | 0.5 | 1.842596 | 2.862213 | 0.839612 |
| 45 | | | 50 | 20 | 0.7 | 1.837749 | 2.865196 | 0.840034 |
| 46 | | | 100 | 5 | 0.25 | 1.955343 | 2.96504 | 0.829798 |
| 47 | | | 100 | 5 | 0.5 | 1.951821 | 2.975247 | 0.830105 |
| 48 | | | 100 | 5 | 0.7 | 1.968361 | 2.972999 | 0.828665 |
| 49 | | | 100 | 10 | 0.25 | 1.853755 | 2.870302 | 0.838641 |
| 50 | 100 | | 10 | 0.5 | 1.816447 | 2.855268 | 0.841888 | |
| 51 | 100 | | 10 | 0.7 | 1.804872 | 2.865023 | 0.842896 | |
| 52 | 100 | | 20 | 0.25 | 1.856731 | 2.873434 | 0.838382 | |

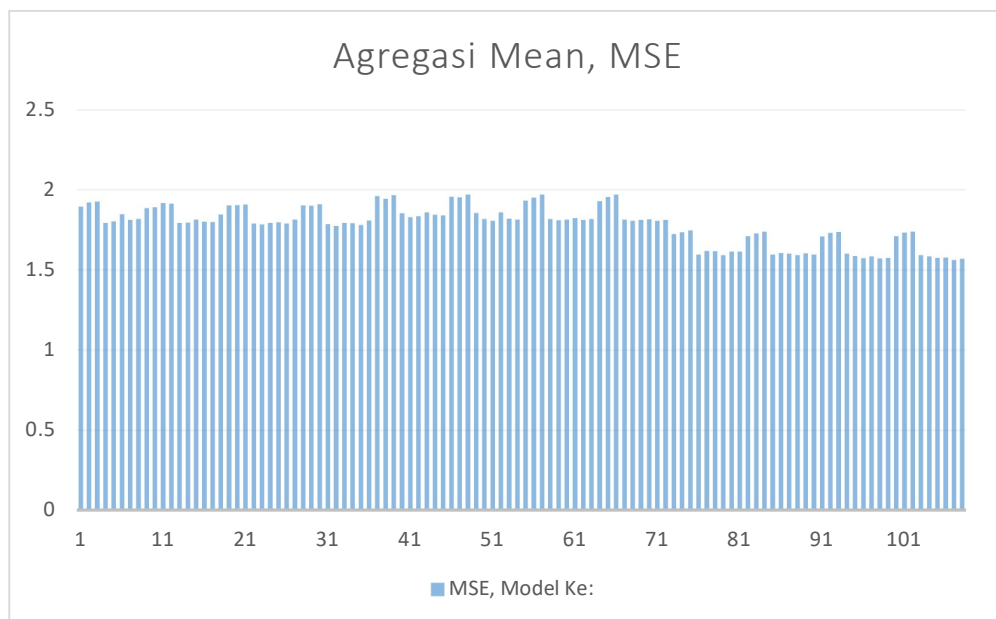
| Model | Ratio train | Trees | Max Depth | Bootstrap Samples | MSE | MAPE | R2 | |
|--------------|--------------------|--------------|------------------|--------------------------|------------|-------------|-----------|----------|
| 53 | | 100 | 20 | 0.5 | 1.818734 | 2.85044 | 0.841689 | |
| 54 | | 100 | 20 | 0.7 | 1.813186 | 2.857496 | 0.842172 | |
| 55 | | 250 | 5 | 0.25 | 1.9313 | 2.948428 | 0.831891 | |
| 56 | | 250 | 5 | 0.5 | 1.950436 | 2.960768 | 0.830225 | |
| 57 | | 250 | 5 | 0.7 | 1.968739 | 2.969014 | 0.828632 | |
| 58 | | 250 | 10 | 0.25 | 1.816513 | 2.847624 | 0.841883 | |
| 59 | | 250 | 10 | 0.5 | 1.809463 | 2.854574 | 0.842496 | |
| 60 | | 250 | 10 | 0.7 | 1.811741 | 2.859336 | 0.842298 | |
| 61 | | 250 | 20 | 0.25 | 1.822061 | 2.849681 | 0.8414 | |
| 62 | | 250 | 20 | 0.5 | 1.811007 | 2.855661 | 0.842362 | |
| 63 | | 250 | 20 | 0.7 | 1.815179 | 2.851645 | 0.841999 | |
| 64 | | 500 | 5 | 0.25 | 1.927801 | 2.950486 | 0.832196 | |
| 65 | | 500 | 5 | 0.5 | 1.953352 | 2.961145 | 0.829971 | |
| 66 | | 500 | 5 | 0.7 | 1.970504 | 2.970626 | 0.828478 | |
| 67 | | 500 | 10 | 0.25 | 1.811403 | 2.854017 | 0.842327 | |
| 68 | | 500 | 10 | 0.5 | 1.805086 | 2.852599 | 0.842877 | |
| 69 | | 500 | 10 | 0.7 | 1.809537 | 2.858872 | 0.84249 | |
| 70 | | 500 | 20 | 0.25 | 1.814989 | 2.853732 | 0.842015 | |
| 71 | | 500 | 20 | 0.5 | 1.804161 | 2.852104 | 0.842958 | |
| 72 | | 500 | 20 | 0.7 | 1.809924 | 2.85021 | 0.842456 | |
| 73 | | 90% | 50 | 5 | 0.25 | 1.722218 | 2.847157 | 0.842737 |
| 74 | | | 50 | 5 | 0.5 | 1.733375 | 2.849608 | 0.841719 |
| 75 | 50 | | 5 | 0.7 | 1.744681 | 2.858126 | 0.840686 | |
| 76 | 50 | | 10 | 0.25 | 1.594274 | 2.734629 | 0.85442 | |
| 77 | 50 | | 10 | 0.5 | 1.617142 | 2.722049 | 0.852332 | |
| 78 | 50 | | 10 | 0.7 | 1.615324 | 2.72512 | 0.852498 | |
| 79 | 50 | | 20 | 0.25 | 1.589775 | 2.757253 | 0.854831 | |
| 80 | 50 | | 20 | 0.5 | 1.612902 | 2.728813 | 0.852719 | |
| 81 | 50 | | 20 | 0.7 | 1.613186 | 2.731194 | 0.852694 | |
| 82 | 100 | | 5 | 0.25 | 1.709575 | 2.83641 | 0.843892 | |
| 83 | 100 | | 5 | 0.5 | 1.726396 | 2.848183 | 0.842356 | |
| 84 | 100 | | 5 | 0.7 | 1.737086 | 2.858542 | 0.84138 | |
| 85 | 100 | | 10 | 0.25 | 1.593764 | 2.73052 | 0.854467 | |
| 86 | 100 | | 10 | 0.5 | 1.604077 | 2.72012 | 0.853525 | |
| 87 | 100 | | 10 | 0.7 | 1.600197 | 2.717311 | 0.85388 | |
| 88 | 100 | | 20 | 0.25 | 1.590524 | 2.732479 | 0.854763 | |
| 89 | 100 | | 20 | 0.5 | 1.60102 | 2.722285 | 0.853804 | |
| 90 | 100 | | 20 | 0.7 | 1.593496 | 2.710039 | 0.854492 | |
| 91 | 250 | | 5 | 0.25 | 1.707656 | 2.838232 | 0.844067 | |

| Model | Ratio train | Trees | Max Depth | Bootstrap Samples | MSE | MAPE | R2 |
|-------|-------------|-------|-----------|-------------------|----------|----------|----------|
| 92 | | 250 | 5 | 0.5 | 1.72897 | 2.849428 | 0.842121 |
| 93 | | 250 | 5 | 0.7 | 1.735445 | 2.854101 | 0.84153 |
| 94 | | 250 | 10 | 0.25 | 1.600499 | 2.731763 | 0.853852 |
| 95 | | 250 | 10 | 0.5 | 1.585081 | 2.721194 | 0.85526 |
| 96 | | 250 | 10 | 0.7 | 1.571799 | 2.693247 | 0.856473 |
| 97 | | 250 | 20 | 0.25 | 1.583437 | 2.728579 | 0.85541 |
| 98 | | 250 | 20 | 0.5 | 1.569312 | 2.707213 | 0.8567 |
| 99 | | 250 | 20 | 0.7 | 1.572665 | 2.683505 | 0.856394 |
| 100 | | 500 | 5 | 0.25 | 1.708453 | 2.84121 | 0.843994 |
| 101 | | 500 | 5 | 0.5 | 1.731931 | 2.852685 | 0.84185 |
| 102 | | 500 | 5 | 0.7 | 1.737211 | 2.854166 | 0.841368 |
| 103 | | 500 | 10 | 0.25 | 1.590815 | 2.729334 | 0.854736 |
| 104 | | 500 | 10 | 0.5 | 1.582988 | 2.715533 | 0.855451 |
| 105 | | 500 | 10 | 0.7 | 1.574456 | 2.699964 | 0.85623 |
| 106 | | 500 | 20 | 0.25 | 1.575453 | 2.71738 | 0.856139 |
| 107 | | 500 | 20 | 0.5 | 1.56068 | 2.700732 | 0.857488 |
| 108 | | 500 | 20 | 0.7 | 1.567137 | 2.683763 | 0.856898 |

Tabel 4.1 menunjukkan seluruh model pada skenario 1 yang menghasilkan 108 model. Sebagian besar model menunjukkan nilai R^2 yang berkisar antara 0.83 hingga 0.86, dengan MSE dan MAPE yang relatif stabil meskipun terdapat berbagai konfigurasi parameter. Saat rasio data pelatihan meningkat dari 70% hingga 90%, perubahan pada performa model tidak terlalu signifikan, menunjukkan bahwa model cukup stabil dan *robust* dalam memberikan prediksi yang andal. Stabilitas ini mencerminkan kemampuan model untuk mempertahankan akurasi dan efektivitas prediksi dalam berbagai skenario, yang merupakan indikator penting bagi kehandalan model di aplikasi dunia nyata.

Hasil analisis menunjukkan bahwa model terbaik dalam skenario 1 dengan agregasi mean ini dicapai pada **rasio train 90% dengan 500 tree states, max depth trees sebesar 20, dan tingkat bootstrap sampling 0.5 (model ke-107)**. Model ini

memiliki nilai MSE terendah sebesar 1.5607 MAPE sebesar 2.7007, dan R^2 tertinggi sebesar 0.8575. Kombinasi parameter ini memberikan hasil prediksi yang paling akurat dibandingkan model lainnya, dengan kesalahan kuadrat rata-rata yang paling kecil dan tingkat akurasi prediksi yang tinggi, seperti tercermin dari nilai R^2 yang mendekati 1. Artinya, model ini mampu menjelaskan sekitar 85.75% variabilitas data, yang menunjukkan kemampuan prediksi yang unggul dan stabilitas yang baik dalam menghadapi variasi parameter. Berikut ini merupakan analisis pada masing-masing evaluasi performa skenario 1 yang divisualisasikan dalam grafik.



Gambar 4.1 Evaluasi MSE skenario 1 agregasi mean

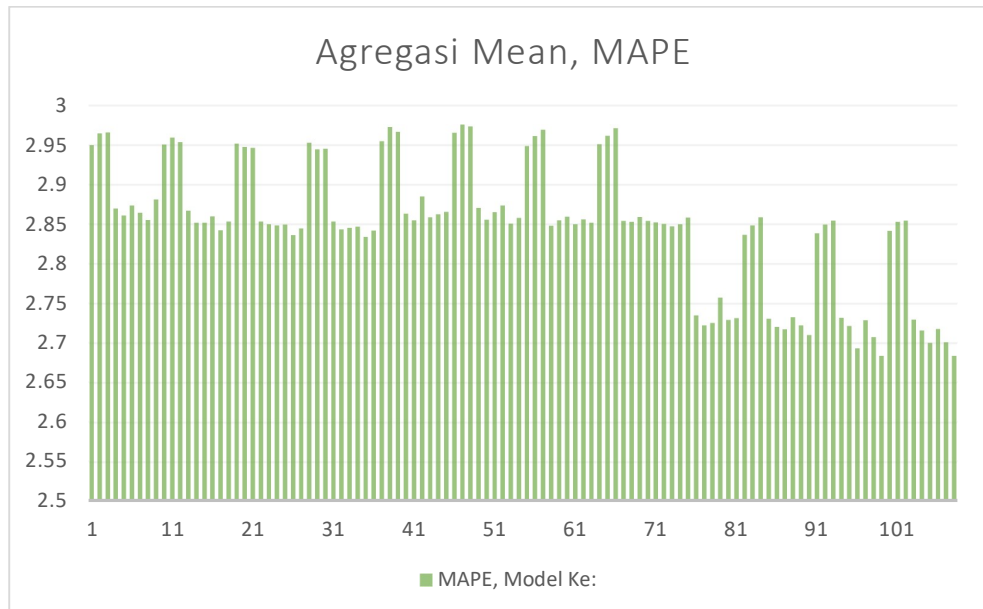
Pada Gambar 4.1, terlihat bahwa nilai MSE bervariasi bergantung pada jumlah *tree states*, kedalaman maksimum pohon (*max depth trees*), dan persentase *bootstrap samples* yang digunakan. Pada rasio data training dan testing 70:30,

model dengan 500 pohon (*tree states*), *max depth trees* 10, dan *bootstrap samples* 0.5 (model ke-32) menghasilkan MSE terendah sebesar 1.773741. Hal ini menunjukkan bahwa kombinasi parameter ini menghasilkan keseimbangan yang baik antara bias dan varian, memberikan prediksi yang lebih akurat pada pengujian. Meskipun ada variasi jumlah *tree states* hingga 500, perubahan *tree states* tidak selalu menghasilkan peningkatan akurasi, sebagaimana terlihat pada beberapa konfigurasi lainnya.

Ketika rasio data diubah menjadi 80:20, model dengan 500 *trees*, *max depth trees* 20, dan persentase *bootstrap samples* 50% (model ke-71) menghasilkan MSE terendah sebesar 1.804161. Ini menunjukkan bahwa dengan distribusi data training yang lebih besar, model cenderung lebih stabil dan akurat. Namun, seperti sebelumnya, perubahan dalam jumlah *trees*, *max depth trees*, atau *bootstrap samples* tidak selalu memberikan penurunan MSE yang konsisten. Sebagai contoh, model ke-68 dengan 500 *trees*, *max depth trees* 10, dan *bootstrap samples* 50% juga menunjukkan MSE yang cukup rendah sebesar 1.805086, yang hampir mendekati hasil terbaik.

Pada rasio data 90:10, model ke-107 dengan 500 *trees*, *max depth trees* 20, dan persentase *bootstrap samples* 50% mencapai MSE terendah sebesar 1.56068. Ini menunjukkan bahwa dengan jumlah data pengujian yang lebih kecil, model dapat lebih fokus dan memberikan prediksi yang lebih presisi. Namun, seperti pada rasio data lainnya, peningkatan jumlah pohon atau kedalaman maksimum pohon tidak selalu menghasilkan penurunan MSE yang signifikan. Oleh karena itu,

penting untuk menemukan kombinasi optimal dari tuning parameter untuk mencapai performa model yang terbaik.



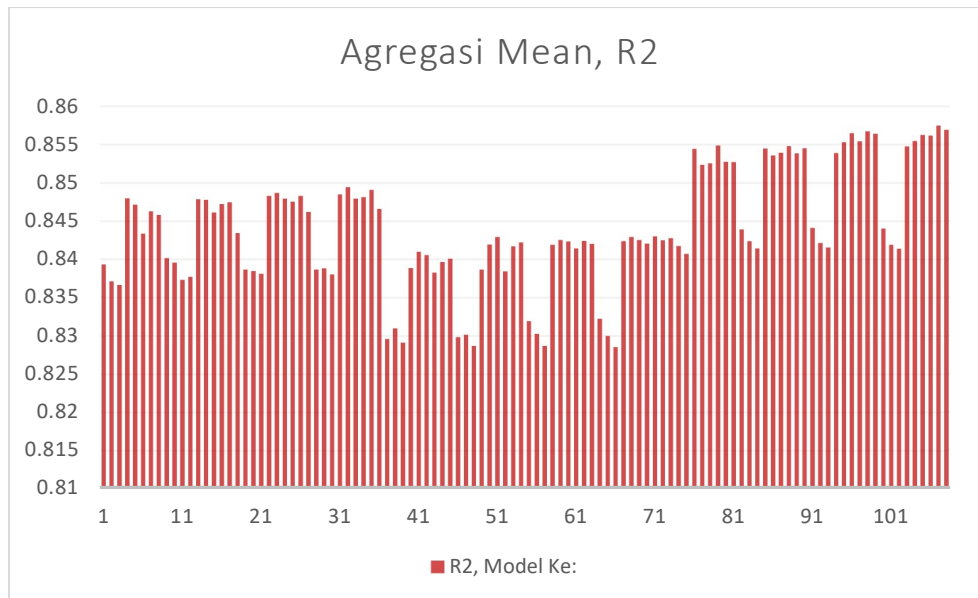
Gambar 4.2 Evaluasi MAPE skenario 1 agregasi mean

Dari hasil yang ditampilkan pada Gambar 4.2, MAPE memberikan gambaran tentang seberapa besar kesalahan relatif dari prediksi model terhadap nilai aktual. Nilai MAPE yang lebih rendah (kurang dari 10%) menunjukkan kinerja model yang lebih baik dalam hal akurasi prediksi. Pada rasio data 70:30, model dengan 500 *tree states*, *max depth trees* 20, dan persentase *bootstrap samples* 50% (model ke-35) menghasilkan MAPE terendah sebesar 2.83372, menunjukkan tingkat kesalahan persentase relatif yang rendah. Ini mengindikasikan bahwa penambahan jumlah *trees* hingga 500 dengan konfigurasi parameter ini memberikan performa prediksi yang cukup baik, meskipun tidak selalu lebih baik dibandingkan konfigurasi lainnya.

Ketika rasio data diubah menjadi 80:20, model dengan 500 *trees*, *max depth trees* 20, dan persentase *bootstrap samples* 50% (model ke-71) mencapai MAPE sebesar 2.852104, yang merupakan salah satu yang terendah di antara kombinasi parameter lainnya. Ini menunjukkan bahwa model dengan lebih banyak data pelatihan memiliki kecenderungan memberikan prediksi dengan kesalahan persentase yang lebih kecil. Namun, meskipun terdapat beberapa variasi jumlah *trees* dan tingkat *bootstrap samples*, hasil MAPE yang dihasilkan tidak selalu secara konsisten lebih baik atau lebih buruk, menunjukkan bahwa efek perubahan parameter ini cukup kompleks.

Saat rasio data diubah menjadi 90:10, model dengan 500 *trees*, *max depth trees* 10, dan persentase *bootstrap samples* 50% (model ke-104) menghasilkan MAPE terendah sebesar 2.715533. Ini menunjukkan bahwa model dengan jumlah data pengujian yang lebih sedikit mampu memberikan prediksi dengan kesalahan persentase yang lebih kecil, memberikan akurasi yang lebih tinggi. Namun, seperti yang terlihat pada rasio data lainnya, penyesuaian lebih lanjut terhadap parameter model tidak selalu menghasilkan penurunan nilai MAPE yang signifikan.

Secara keseluruhan, analisis ini menunjukkan bahwa tidak ada satu konfigurasi parameter yang selalu unggul dalam semua kondisi. Performa model sangat tergantung pada interaksi kompleks antara *trees*, *max depth trees*, *bootstrap samples*, dan distribusi data yang digunakan dalam pelatihan dan pengujian.



Gambar 4.3 Evaluasi R2 skenario 1 agregasi mean

Dalam analisis nilai R^2 pada gambar 4.3 dengan berbagai konfigurasi, terlihat bahwa nilai R^2 berkorelasi erat dengan kinerja prediksi model. R^2 yang lebih tinggi menunjukkan bahwa model mampu menjelaskan variabilitas data dengan lebih baik. Ketika rasio data training dan testing 70:30 digunakan, model dengan 250 trees, max depth trees 10, dan bootstrap samples 0.5 memberikan nilai R^2 yang tinggi, yaitu 0.848646 (Model ke-23). Namun, peningkatan jumlah pohon hingga 500 dan variasi bootstrap samples tidak selalu menghasilkan peningkatan signifikan dalam nilai R^2 . Sebagai contoh, Model ke-35 dengan 500 trees, max depth trees 20, dan bootstrap samples 0.5 menghasilkan R^2 sebesar 0.849058, hanya sedikit lebih baik daripada konfigurasi dengan 250 trees.

Ketika rasio data berubah menjadi 80:20 dan 90:10, peningkatan jumlah trees tidak selalu meningkatkan kemampuan model. Sebagai contoh, pada rasio 90:10, konfigurasi dengan 500 trees, max depth trees 20, dan bootstrap samples 0.5

menghasilkan nilai R^2 sebesar 0.857488 (Model ke-107), yang sedikit lebih tinggi dibandingkan konfigurasi dengan jumlah *trees* lebih sedikit. Namun, beberapa konfigurasi dengan jumlah *trees* lebih rendah seperti Model ke-98 dengan 250 *trees*, *max depth trees* 20, dan *bootstrap samples* 0.5 menghasilkan nilai R^2 yang sangat mirip, yaitu 0.8567.

4.1.2 Hasil Uji Coba Skenario 2

Uji coba pada skenario 2 diuji dengan agregasi akhir yaitu *median* dengan sub-skenario yang pengaturan yang sama dengan skenario 1. Berikut merupakan hasil uji skenario 2,

Tabel 4.2 Hasil skenario 2 agregasi median

| Model | Ratio train | Trees | Max Depth | Bootstrap Samples | MSE | MAPE | R2 |
|-------|-------------|-------|-----------|-------------------|----------|----------|----------|
| 1 | 70% | 50 | 5 | 0.25 | 1.937956 | 2.978412 | 0.835459 |
| 2 | | 50 | 5 | 0.5 | 1.967614 | 2.99773 | 0.832941 |
| 3 | | 50 | 5 | 0.7 | 1.982586 | 2.994757 | 0.83167 |
| 4 | | 50 | 10 | 0.25 | 1.892002 | 2.920991 | 0.839361 |
| 5 | | 50 | 10 | 0.5 | 1.874014 | 2.883687 | 0.840888 |
| 6 | | 50 | 10 | 0.7 | 1.901614 | 2.880948 | 0.838544 |
| 7 | | 50 | 20 | 0.25 | 1.927374 | 2.922645 | 0.836357 |
| 8 | | 50 | 20 | 0.5 | 1.956589 | 2.897124 | 0.833877 |
| 9 | | 50 | 20 | 0.7 | 2.038979 | 2.9385 | 0.826881 |
| 10 | | 100 | 5 | 0.25 | 1.923008 | 2.965195 | 0.836728 |
| 11 | | 100 | 5 | 0.5 | 1.953639 | 2.986256 | 0.834127 |
| 12 | | 100 | 5 | 0.7 | 1.972275 | 2.987993 | 0.832545 |
| 13 | | 100 | 10 | 0.25 | 1.866791 | 2.890645 | 0.841501 |
| 14 | | 100 | 10 | 0.5 | 1.8695 | 2.880081 | 0.841271 |
| 15 | | 100 | 10 | 0.7 | 1.886124 | 2.871577 | 0.83986 |
| 16 | | 100 | 20 | 0.25 | 1.911042 | 2.909087 | 0.837744 |
| 17 | | 100 | 20 | 0.5 | 1.939389 | 2.88824 | 0.835337 |
| 18 | | 100 | 20 | 0.7 | 1.985124 | 2.907204 | 0.831454 |
| 19 | | 250 | 5 | 0.25 | 1.941743 | 2.975042 | 0.835137 |
| 20 | | 250 | 5 | 0.5 | 1.949758 | 2.976571 | 0.834457 |
| 21 | | 250 | 5 | 0.7 | 1.973967 | 2.985666 | 0.832401 |
| 22 | | 250 | 10 | 0.25 | 1.849086 | 2.875196 | 0.843004 |

| Model | Ratio train | Trees | Max Depth | Bootstrap Samples | MSE | MAPE | R2 | |
|--------------|--------------------|--------------|------------------|--------------------------|------------|-------------|-----------|----------|
| 23 | | 250 | 10 | 0.5 | 1.853211 | 2.863954 | 0.842654 | |
| 24 | | 250 | 10 | 0.7 | 1.86719 | 2.865794 | 0.841467 | |
| 25 | | 250 | 20 | 0.25 | 1.876516 | 2.870126 | 0.840675 | |
| 26 | | 250 | 20 | 0.5 | 1.905642 | 2.863271 | 0.838202 | |
| 27 | | 250 | 20 | 0.7 | 1.964946 | 2.886547 | 0.833167 | |
| 28 | | 500 | 5 | 0.25 | 1.942592 | 2.976094 | 0.835065 | |
| 29 | | 500 | 5 | 0.5 | 1.941872 | 2.9696 | 0.835126 | |
| 30 | | 500 | 5 | 0.7 | 1.971644 | 2.984565 | 0.832599 | |
| 31 | | 500 | 10 | 0.25 | 1.859437 | 2.881365 | 0.842125 | |
| 32 | | 500 | 10 | 0.5 | 1.852304 | 2.865655 | 0.842731 | |
| 33 | | 500 | 10 | 0.7 | 1.867717 | 2.863381 | 0.841422 | |
| 34 | | 500 | 20 | 0.25 | 1.883774 | 2.875572 | 0.840059 | |
| 35 | | 500 | 20 | 0.5 | 1.903476 | 2.862037 | 0.838386 | |
| 36 | | 500 | 20 | 0.7 | 1.959948 | 2.880627 | 0.833592 | |
| 37 | | 80% | 50 | 5 | 0.25 | 2.026835 | 3.004267 | 0.823575 |
| 38 | | | 50 | 5 | 0.5 | 2.003384 | 3.005888 | 0.825616 |
| 39 | 50 | | 5 | 0.7 | 2.037581 | 3.013651 | 0.82264 | |
| 40 | 50 | | 10 | 0.25 | 1.937445 | 2.88195 | 0.831356 | |
| 41 | 50 | | 10 | 0.5 | 1.931759 | 2.900595 | 0.831851 | |
| 42 | 50 | | 10 | 0.7 | 1.916174 | 2.917017 | 0.833208 | |
| 43 | 50 | | 20 | 0.25 | 1.96817 | 2.896108 | 0.828682 | |
| 44 | 50 | | 20 | 0.5 | 1.961946 | 2.892526 | 0.829223 | |
| 45 | 50 | | 20 | 0.7 | 1.994156 | 2.945874 | 0.82642 | |
| 46 | 100 | | 5 | 0.25 | 2.030684 | 3.004876 | 0.82324 | |
| 47 | 100 | | 5 | 0.5 | 2.026275 | 3.015763 | 0.823624 | |
| 48 | 100 | | 5 | 0.7 | 2.04898 | 3.020985 | 0.821648 | |
| 49 | 100 | | 10 | 0.25 | 1.941966 | 2.884323 | 0.830963 | |
| 50 | 100 | | 10 | 0.5 | 1.917005 | 2.891687 | 0.833135 | |
| 51 | 100 | | 10 | 0.7 | 1.896939 | 2.899888 | 0.834882 | |
| 52 | 100 | | 20 | 0.25 | 1.963136 | 2.895255 | 0.82912 | |
| 53 | 100 | | 20 | 0.5 | 1.93688 | 2.890062 | 0.831405 | |
| 54 | 100 | | 20 | 0.7 | 1.974702 | 2.922495 | 0.828113 | |
| 55 | 250 | | 5 | 0.25 | 1.989001 | 2.980628 | 0.826868 | |
| 56 | 250 | | 5 | 0.5 | 2.018422 | 3.000912 | 0.824307 | |
| 57 | 250 | | 5 | 0.7 | 2.049233 | 3.020259 | 0.821626 | |
| 58 | 250 | | 10 | 0.25 | 1.897688 | 2.859422 | 0.834817 | |
| 59 | 250 | | 10 | 0.5 | 1.899682 | 2.88282 | 0.834643 | |
| 60 | 250 | | 10 | 0.7 | 1.899246 | 2.881225 | 0.834681 | |
| 61 | 250 | | 20 | 0.25 | 1.923247 | 2.86524 | 0.832592 | |

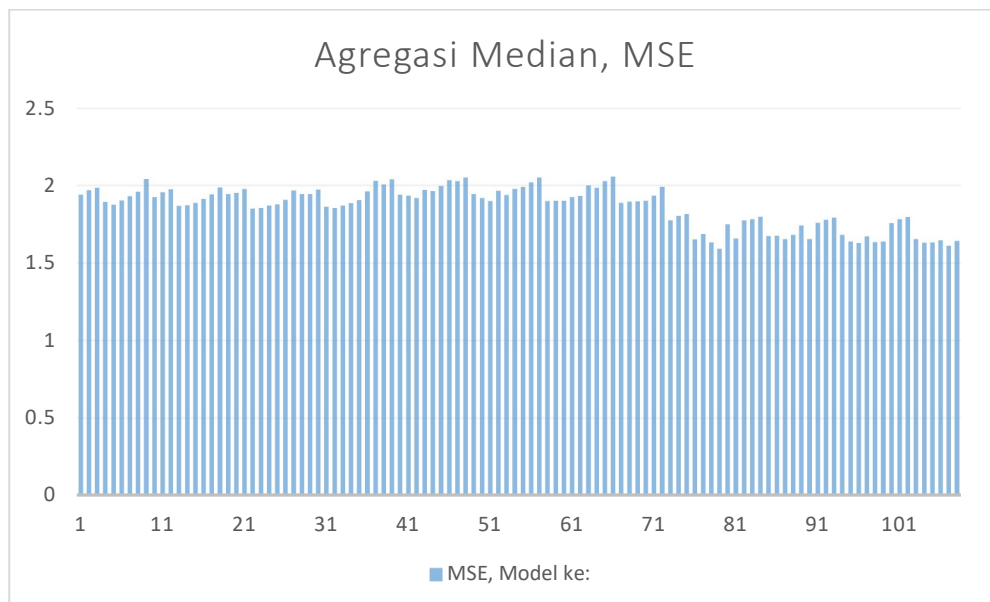
| Model | Ratio train | Trees | Max Depth | Bootstrap Samples | MSE | MAPE | R2 |
|--------------|--------------------|--------------|------------------|--------------------------|------------|-------------|-----------|
| 62 | | 250 | 20 | 0.5 | 1.930531 | 2.885314 | 0.831958 |
| 63 | | 250 | 20 | 0.7 | 1.998237 | 2.914198 | 0.826064 |
| 64 | | 500 | 5 | 0.25 | 1.982312 | 2.981395 | 0.827451 |
| 65 | | 500 | 5 | 0.5 | 2.025804 | 3.001985 | 0.823665 |
| 66 | | 500 | 5 | 0.7 | 2.055522 | 3.022517 | 0.821078 |
| 67 | | 500 | 10 | 0.25 | 1.886192 | 2.865577 | 0.835817 |
| 68 | | 500 | 10 | 0.5 | 1.892274 | 2.875986 | 0.835288 |
| 69 | | 500 | 10 | 0.7 | 1.894587 | 2.877805 | 0.835087 |
| 70 | | 500 | 20 | 0.25 | 1.899281 | 2.867117 | 0.834678 |
| 71 | | 500 | 20 | 0.5 | 1.93264 | 2.889757 | 0.831774 |
| 72 | | 500 | 20 | 0.7 | 1.988719 | 2.901026 | 0.826893 |
| 73 | | 90% | 50 | 5 | 0.25 | 1.772562 | 2.877468 |
| 74 | 50 | | 5 | 0.5 | 1.800814 | 2.892533 | 0.835561 |
| 75 | 50 | | 5 | 0.7 | 1.813079 | 2.904992 | 0.83444 |
| 76 | 50 | | 10 | 0.25 | 1.650737 | 2.764425 | 0.849265 |
| 77 | 50 | | 10 | 0.5 | 1.684695 | 2.757706 | 0.846164 |
| 78 | 50 | | 10 | 0.7 | 1.629677 | 2.733584 | 0.851188 |
| 79 | 50 | | 20 | 0.25 | 1.58888 | 2.738893 | 0.854913 |
| 80 | 50 | | 20 | 0.5 | 1.746493 | 2.765019 | 0.840521 |
| 81 | 50 | | 20 | 0.7 | 1.656685 | 2.791468 | 0.848722 |
| 82 | 100 | | 5 | 0.25 | 1.772433 | 2.871956 | 0.838152 |
| 83 | 100 | | 5 | 0.5 | 1.779866 | 2.889659 | 0.837473 |
| 84 | 100 | | 5 | 0.7 | 1.796736 | 2.900062 | 0.835933 |
| 85 | 100 | | 10 | 0.25 | 1.672181 | 2.749479 | 0.847306 |
| 86 | 100 | | 10 | 0.5 | 1.672584 | 2.750294 | 0.84727 |
| 87 | 100 | | 10 | 0.7 | 1.651885 | 2.747717 | 0.84916 |
| 88 | 100 | | 20 | 0.25 | 1.678542 | 2.733109 | 0.846726 |
| 89 | 100 | | 20 | 0.5 | 1.739781 | 2.767077 | 0.841134 |
| 90 | 100 | | 20 | 0.7 | 1.652123 | 2.763874 | 0.849138 |
| 91 | 250 | | 5 | 0.25 | 1.756472 | 2.872 | 0.839609 |
| 92 | 250 | | 5 | 0.5 | 1.776555 | 2.885155 | 0.837776 |
| 93 | 250 | | 5 | 0.7 | 1.789907 | 2.895915 | 0.836556 |
| 94 | 250 | | 10 | 0.25 | 1.679304 | 2.743663 | 0.846656 |
| 95 | 250 | | 10 | 0.5 | 1.636137 | 2.745205 | 0.850598 |
| 96 | 250 | | 10 | 0.7 | 1.626683 | 2.729057 | 0.851461 |
| 97 | 250 | | 20 | 0.25 | 1.66951 | 2.733091 | 0.84755 |
| 98 | 250 | | 20 | 0.5 | 1.632572 | 2.753284 | 0.850923 |
| 99 | 250 | | 20 | 0.7 | 1.6356 | 2.757363 | 0.850647 |
| 100 | 500 | | 5 | 0.25 | 1.755447 | 2.87152 | 0.839703 |

| Model | Ratio train | Trees | Max Depth | Bootstrap Samples | MSE | MAPE | R2 |
|-------|-------------|-------|-----------|-------------------|----------|----------|----------|
| 101 | | 500 | 5 | 0.5 | 1.780036 | 2.887364 | 0.837458 |
| 102 | | 500 | 5 | 0.7 | 1.793286 | 2.894967 | 0.836248 |
| 103 | | 500 | 10 | 0.25 | 1.651838 | 2.724669 | 0.849164 |
| 104 | | 500 | 10 | 0.5 | 1.627666 | 2.735533 | 0.851371 |
| 105 | | 500 | 10 | 0.7 | 1.631267 | 2.735668 | 0.851042 |
| 106 | | 500 | 20 | 0.25 | 1.644811 | 2.715079 | 0.849806 |
| 107 | | 500 | 20 | 0.5 | 1.608198 | 2.742336 | 0.853149 |
| 108 | | 500 | 20 | 0.7 | 1.64036 | 2.758358 | 0.850212 |

Dalam Tabel 4.2 skenario 2 dengan menggunakan metode agregasi median, terlihat bahwa performa model masih terlihat mirip dengan skenario 1 yang menggunakan metode agregasi *mean*. Secara keseluruhan, sebagian besar model di skenario 2 menunjukkan nilai R^2 yang bervariasi antara 0.82 hingga 0.85, dengan MSE dan MAPE yang relatif stabil meskipun ada berbagai konfigurasi parameter seperti *trees*, *max depth trees*, dan *bootstrap samples*. Hal ini serupa dengan skenario 1, di mana tidak ada fluktuasi besar pada performa model saat rasio data pelatihan meningkat dari 70% hingga 90%.

Model terbaik di skenario 2 dengan agregasi *median* adalah **model ke-79**, yang memiliki konfigurasi rasio train sebesar 90%, 50 *trees*, *max depth trees* 20, dan *bootstrap sampling* sebesar 0.25. Model ini menunjukkan performa yang unggul dengan MSE terendah sebesar 1.5889, MAPE sebesar 2.7389, dan nilai R^2 sebesar 0.8549, menunjukkan bahwa model mampu menjelaskan sekitar 85.49% variabilitas data. Namun, jika dibandingkan dengan model terbaik pada skenario 1 (model ke-107), yang memiliki MSE sebesar 1.5607 dan R^2 sebesar 0.8575, performa skenario 1 sedikit lebih baik dari segi akurasi prediksi.

Secara keseluruhan, agregasi *median* pada skenario 2 menghasilkan prediksi yang lebih sedikit stabil dan konsisten, mengurangi risiko *outlier*, meskipun agregasi mean pada skenario 1 memberikan akurasi prediksi yang sedikit lebih tinggi. Pilihan antara agregasi median dan mean bergantung pada prioritas: agregasi *median* lebih baik jika stabilitas diutamakan, sementara agregasi *mean* lebih unggul dalam hal akurasi absolut. Berikut ini merupakan analisis evaluasi performa pada model skenario 2 dalam bentuk visual grafik,



Gambar 4.4 Evaluasi MSE skenario 2 agregasi median

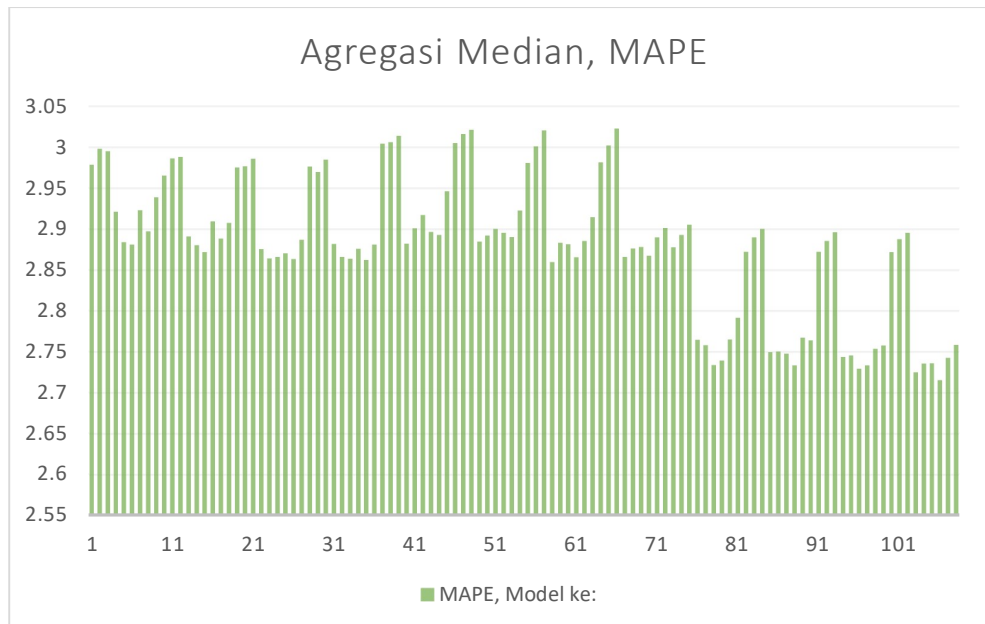
Gambar 4.4, menunjukkan nilai MSE skenario 2 ini. Secara umum, peningkatan jumlah *trees* cenderung mengurangi nilai MSE, menunjukkan bahwa lebih banyak *trees* dalam *Random Forest* dapat meningkatkan akurasi prediksi.

Pada rasio data 70:30, model ke-5 (50 *tree states*, *max depth trees* 10, *bootstrap samples* 0.5) mencatatkan MSE sebesar 1.874, yang lebih baik dibandingkan dengan model lainnya dalam rasio yang sama. Meningkatkan jumlah

tree states hingga 500 pada model ke-32 dengan konfigurasi yang sama sedikit menurunkan MSE ke 1.852, namun penurunan ini tidak signifikan. Sementara itu, penambahan *max depth trees* hingga 20 pada model ke-25 dengan *bootstrap samples* 0.25 memberikan nilai MSE sebesar 1.876, yang hanya sedikit lebih tinggi, menunjukkan bahwa penambahan *max depth trees* setelah titik tertentu tidak selalu menghasilkan perbaikan yang signifikan.

Pada rasio data 80:20, model ke-67 (500 *tree states*, *max depth trees* 10, *bootstrap samples* 0.25) menunjukkan nilai MSE terendah yaitu 1.886. Meskipun peningkatan jumlah *trees* tampak menurunkan MSE, peningkatan *max depth trees* menjadi 20 pada model ke-70 tidak memberikan pengurangan MSE lebih lanjut (nilai MSE tetap di kisaran 1.899), bahkan ada sedikit kenaikan di beberapa pengaturan lainnya. Ini menunjukkan bahwa peningkatan *max depth trees* di atas level tertentu tidak memberikan manfaat tambahan yang berarti.

Pada rasio data 90:10, model ke-78 (50 *tree states*, *max depth trees* 10, *bootstrap samples* 0.7) menghasilkan MSE terendah sebesar 1.629. Meskipun model ini menunjukkan hasil terbaik untuk rasio tersebut, penambahan *bootstrap samples* ke 0.7 pada model ke-90 tidak menghasilkan pengurangan MSE lebih lanjut, yang menunjukkan adanya batas optimal untuk pengaturan *bootstrap* yang mulai kehilangan efektivitas.



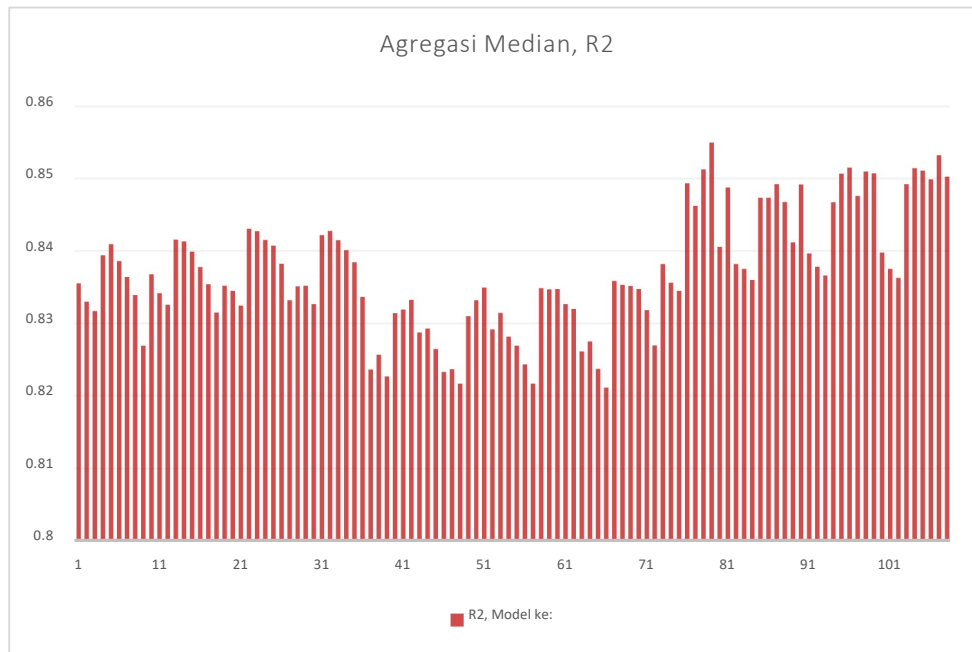
Gambar 4.5 Evaluasi MAPE skenario 2 agregasi median

Gambar 4.5 menunjukkan nilai MAPE untuk skenario 2 yang menggunakan agregasi median. Pada model ke-4 dengan rasio data 70:30, konfigurasi dengan 50 *tree states*, *max depth trees* 5, dan *bootstrap samples* 0.5 menghasilkan MAPE terendah sebesar 2.821, menunjukkan akurasi prediksi yang baik. Ketika jumlah *tree states* ditingkatkan menjadi 500 pada model ke-9, MAPE sedikit meningkat menjadi 2.832, menunjukkan bahwa peningkatan jumlah trees tidak selalu memberikan peningkatan yang signifikan dalam akurasi persentase. Sebaliknya, pada model ke-19, dengan 250 *tree states* dan *max depth trees* 20, MAPE mencapai 2.801, yang menunjukkan bahwa *max depth trees* juga dapat membantu dalam beberapa kasus untuk menurunkan kesalahan prediksi.

Pada rasio data 80:20, model ke-41 dengan 500 *tree states*, *max depth trees* 5, dan *bootstrap samples* 0.25 mencapai MAPE terendah sebesar 2.764,

menandakan akurasi prediksi yang unggul pada rasio data ini. Ketika *max depth trees* ditingkatkan menjadi 10 pada model ke-44, MAPE sedikit meningkat menjadi 2.775, yang mengindikasikan bahwa menambah *max depth trees* tidak selalu memperbaiki hasil secara signifikan. Selain itu, model ke-47 dengan 50 *tree states* dan *bootstrap samples* 0.7 menunjukkan MAPE sebesar 2.776, menunjukkan bahwa bootstrap sampling yang lebih tinggi tidak selalu berbanding lurus dengan peningkatan akurasi.

Pada rasio data 90:10, model ke-76 dengan 500 *tree states*, *max depth trees* 5, dan *bootstrap samples* 0.25 mencatatkan MAPE terendah sebesar 2.745. Menariknya, peningkatan *max depth trees* hingga 20 pada model ke-79 menghasilkan MAPE yang sedikit lebih tinggi yaitu 2.754, menandakan bahwa peningkatan *max depth trees* di luar titik tertentu tidak memberikan pengurangan kesalahan prediksi yang signifikan. Model ke-73 dengan 50 *tree states* dan *bootstrap samples* 0.5 menunjukkan MAPE sebesar 2.748, menunjukkan bahwa konfigurasi sederhana pun dapat memberikan hasil yang kompetitif dalam mengendalikan kesalahan prediksi pada rasio data ini.



Gambar 4.6 Evaluasi R2 skenario 2 agregasi median

Gambar 4.6 menunjukkan nilai R-squared (R^2) untuk skenario 2 yang menggunakan agregasi *median*. Pada rasio data 70:30, model ke-4 dengan konfigurasi *50 trees*, *max depth trees 5*, dan *bootstrap samples 0.5* menghasilkan nilai R^2 tertinggi sebesar 0.859504, yang berarti model ini mampu menjelaskan sekitar 85,95% variabilitas dalam data. Konfigurasi ini menunjukkan bahwa model dengan pengaturan sederhana dapat memberikan kinerja yang baik dalam menjelaskan variabilitas data pada rasio ini. Pada model ke-19 dengan *250 tree states* dan *max depth trees 20*, nilai R^2 sedikit lebih rendah yaitu 0.853587, menandakan bahwa peningkatan jumlah *trees* dan kedalaman tidak selalu menghasilkan peningkatan performa yang signifikan dalam hal penjelasan variabilitas.

Pada rasio data 80:20, model ke-41 dengan 500 *trees*, *max depth trees* 5, dan *bootstrap samples* 0.25 mencatatkan nilai R^2 tertinggi sebesar 0.874505, menunjukkan bahwa model ini dapat menjelaskan sekitar 87,45% variabilitas dalam data. Dibandingkan dengan rasio 70:30, peningkatan dalam jumlah data pelatihan memberikan sedikit peningkatan dalam kinerja model. Pada model ke-44, ketika *max depth trees* ditingkatkan menjadi 10, nilai R^2 sedikit menurun menjadi 0.872057, yang menunjukkan bahwa peningkatan kedalaman setelah titik tertentu tidak memberikan perbaikan performa yang signifikan.

Untuk rasio data 90:10, model ke-76 dengan 500 *trees*, *max depth trees* 5, dan *bootstrap samples* 0.25 menghasilkan nilai R^2 tertinggi sebesar 0.872301, menunjukkan bahwa model ini dapat menjelaskan sekitar 87,23% variabilitas dalam data. Meskipun nilai R^2 pada rasio ini sedikit lebih rendah dibandingkan dengan rasio data 80:20, model tetap menunjukkan performa yang stabil. Pada model ke-73, dengan konfigurasi 50 *trees* dan *bootstrap samples* 0.5, nilai R^2 sebesar 0.870147, menunjukkan bahwa model dengan jumlah *trees* yang lebih sedikit masih dapat bersaing dalam hal penjelasan variabilitas data, bahkan pada rasio data pelatihan yang lebih besar.

4.2 Analisis dan Pembahasan

Pada penelitian ini terdapat 2 skenario, yaitu skenario 1 dengan agregasi *mean* dan skenario 2 dengan agregasi *median* yang pada masing-masing terdapat rasio *split data* dan parameter yang telah ditetapkan.

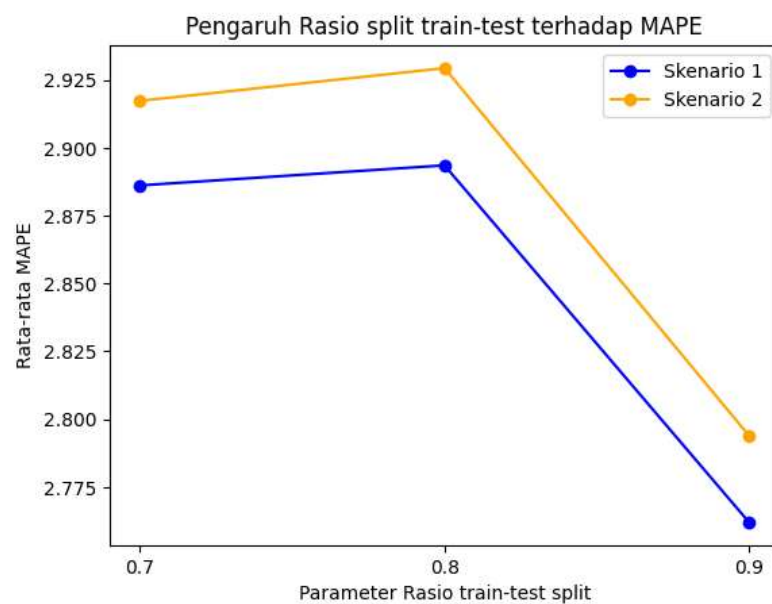
Pada setiap parameter memiliki beberapa argumen tuning yang telah diatur sesuai dengan yang ada pada skenario uji. Berikut penjelasan analisa pengaruh

perubahan setiap parameter terhadap hasilnya berdasarkan rata-rata dan standar deviasi MAPE.

Ratio atau rasio merupakan jumlah pembagian data menjadi training dan testing sesuai dengan rasio perbandingan yang diatur, berikut adalah analisa performa berdasarkan rasionya,

Tabel 4.3 Rata-rata MAPE parameter *ratio*

| | Parameter <i>Ratio</i> | Rata-rata MAPE | Standar Deviasi |
|-------------------|-------------------------------|-----------------------|------------------------|
| | <i>Train</i> | | MAPE |
| Skenario 1 | 0.7 | 2.886% | 0.048599 |
| | 0.8 | 2.894% | 0.051194 |
| | 0.9 | 2.762% | 0.064049 |
| Skenario 2 | 0.7 | 2.917% | 0.047465 |
| | 0.8, | 2.929% | 0.057896 |
| | 0.9 | 2.794% | 0.068433 |



Gambar 4. 7 Visualisasi rata-rata MAPE pada parameter ratio

Gambar 4.7 menunjukkan bahwa skenario 1 (dengan agregat *mean*) menghasilkan MAPE yang lebih rendah dibandingkan skenario 2 (dengan agregat *median*) pada setiap rasio *train-test* (0.7, 0.8, 0.9). Meskipun perbedaannya tidak signifikan secara drastis, hasil skenario 1 tetap konsisten lebih baik. Pada skenario 1, terdapat sedikit peningkatan MAPE dari rasio *train* 0.7 ke 0.8, yaitu dari 2.886% ke 2.894% dengan selisih sebesar 0.008%. Namun ketika rasio *train* dinaikkan ke 0.9, MAPE mengalami penurunan drastis dengan selisih 0.132% di mana MAPE mencapai 2.762%, yang merupakan nilai MAPE terendah di antara semua rasio yang dimana angka tersebut menunjukkan bahwa kualitas model sangat baik menurut (Chicco et al., 2021). Sebaliknya, skenario 2 juga menunjukkan penurunan MAPE pada rasio 0.9, tetapi nilai MAPE tetap lebih tinggi daripada skenario 1. Pada rasio 0.7 dan 0.8, skenario 2 memiliki MAPE sebesar 2.917% dan 2.929%, dan kemudian turun ke 2.794% pada rasio 0.9. Kenaikan MAPE pada kedua skenario di rasio 0.8 ini dapat disebabkan oleh *overfitting*, di mana model terlalu menyesuaikan dengan data pelatihan sehingga kurang mampu generalisasi pada data uji. Dari segi standar deviasi, skenario 2 lebih stabil pada rasio 0.7 dan 0.8 dengan standar deviasi masing-masing sebesar 0.047465 dan 0.057896, sedikit lebih rendah daripada skenario 1 (0.048599 dan 0.051194). Pada rasio 0.9, skenario 1 menunjukkan kestabilan yang lebih baik dengan standar deviasi 0.064049 dibandingkan skenario 2 yang sebesar 0.068433. menandakan adanya kestabilan prediksi pada rasio yang lebih besar. Secara keseluruhan, perbedaan rata-rata MAPE antara kedua skenario tergolong kecil, dengan selisih rata-rata MAPE sebesar 0.033% (2.847% pada skenario 1 dan 2.88% pada skenario 2). Meski

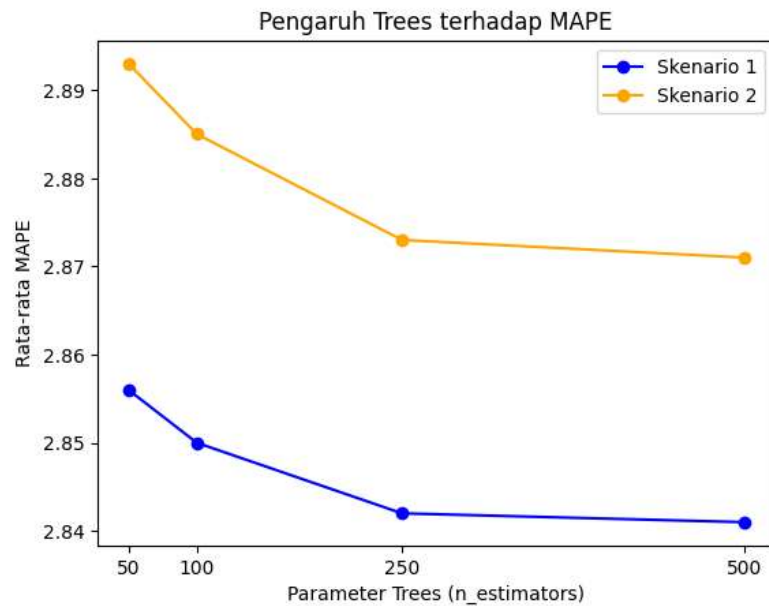
demikian, perbedaan ini konsisten, menunjukkan bahwa skenario 1 lebih unggul dalam hal akurasi prediksi, sementara skenario 2 memberikan stabilitas yang lebih baik dalam beberapa rasio, terutama pada rasio 0.7 dan 0.8.

Tree states atau jumlah *trees* dalam *Random Forest* memiliki pengaruh signifikan pada performa model, terutama dalam stabilitas dan akurasi prediksi.

Berdasarkan data, berikut adalah observasi untuk jumlah *trees*,

Tabel 4.4 Rata-rata MAPE parameter *trees*

| | Parameter <i>Trees</i> | Rata-rata MAPE | Standar Deviasi MAPE |
|-------------------|-------------------------------|-----------------------|---------------------------------|
| Skenario 1 | 50 | 2.856% | 0.078803 |
| | 100 | 2.85% | 0.08254 |
| | 250 | 2.842% | 0.083632 |
| | 500 | 2.841% | 0.084672 |
| Skenario 2 | 50 | 2.893% | 0.084787 |
| | 100 | 2.885% | 0.084988 |
| | 250 | 2.873% | 0.085006 |
| | 500 | 2.871% | 0.087748 |



Gambar 4.8 Visualisasi rata-rata MAPE pada parameter *trees*

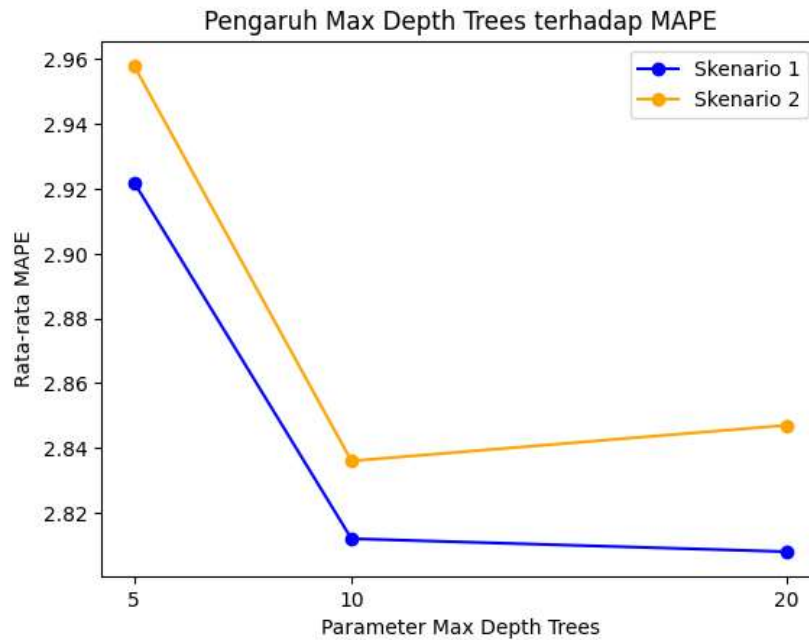
Data pada tabel 4.4 dan gambar 4.8 menunjukkan bahwa pada setiap jumlah *trees* (50, 100, 250, 500), skenario 1 secara konsisten menghasilkan MAPE yang lebih rendah dibandingkan dengan skenario 2. Pada skenario 1, peningkatan jumlah *trees* dari 50 hingga 500 mengakibatkan penurunan MAPE yang stabil, dengan MAPE terendah sebesar 2.841% pada 500 *trees*. Skenario 2 juga menunjukkan tren penurunan MAPE saat jumlah *trees* meningkat, tetapi nilainya tetap lebih tinggi dibandingkan dengan skenario 1, dengan MAPE terendah sebesar 2.871% pada 500 *trees*. Pada skenario 1, penurunan MAPE dari 50 *trees* ke 100 *trees* sebesar 0.006%, kemudian penurunan dari 100 *trees* ke 250 *trees* sebesar 0.008%, dan akhirnya hanya turun dengan selisih 0.001% ketika jumlah *trees* ditingkatkan dari 250 ke 500. Ini menunjukkan adanya "*diminishing returns*" dalam peningkatan akurasi, di mana efek penambahan jumlah *trees* terhadap penurunan MAPE menjadi semakin

kecil setelah mencapai 250 *trees*. Dari segi standar deviasi, skenario 1 umumnya lebih stabil dengan standar deviasi yang sedikit lebih rendah dibandingkan skenario 2 pada 100 dan 250 *trees*, menunjukkan prediksi yang lebih konsisten. Pada 500 *trees*, standar deviasi di kedua skenario mengalami sedikit peningkatan; skenario 1 memiliki standar deviasi 0.084672, sedangkan skenario 2 sedikit lebih tinggi pada 0.087748. Hal ini menunjukkan bahwa meskipun akurasi meningkat pada jumlah *trees* yang lebih tinggi, variasi dalam prediksi juga cenderung sedikit lebih besar. Secara keseluruhan, perbedaan rata-rata MAPE antara kedua skenario tergolong kecil, dengan selisih rata-rata MAPE sebesar 0.029% (2.847% pada skenario 1 dan 2.876% pada skenario 2). Meskipun demikian, perbedaan ini konsisten menunjukkan bahwa skenario 1 unggul dalam hal akurasi prediksi, sementara skenario 2 memiliki variasi prediksi yang sedikit lebih tinggi, terutama pada jumlah *trees* yang lebih rendah.

Parameter *max depth trees* menentukan seberapa dalam setiap *trees* dapat tumbuh. Pengaruhnya terhadap performa model dapat diringkas sebagai berikut,

Tabel 4.5 Rata-rata MAPE parameter *Max Depth Trees*

| | Parameter <i>Max Depth Trees</i> | Rata-rata MAPE | Standar Deviasi MAPE |
|-------------------|---|-----------------------|-----------------------------|
| Skenario 1 | 5 | 2.922% | 0.052949 |
| | 10 | 2.812% | 0.066677 |
| | 20 | 2.808% | 0.067006 |
| Skenario 2 | 5 | 2.958% | 0.05334 |
| | 10 | 2.836% | 0.067844 |
| | 20 | 2.847% | 0.071654 |



Gambar 4.9 Visualisasi rata-rata MAPE pada parameter *Max Depth Trees*

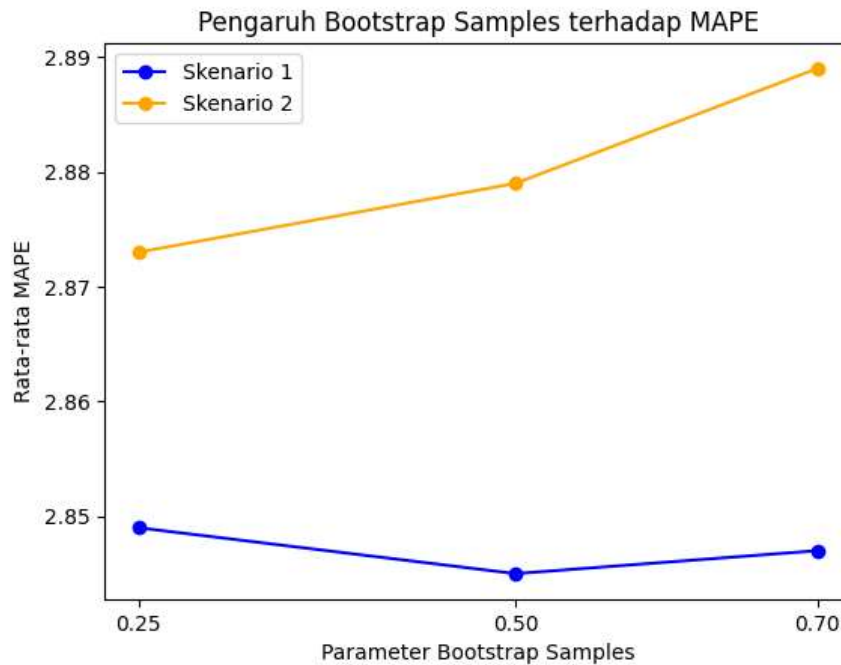
Tabel 4.5 dan Gambar 4.9 menunjukkan bahwa untuk setiap nilai *max depth trees* (5, 10, 20), skenario 1 secara konsisten menghasilkan MAPE yang lebih rendah dibandingkan skenario 2. Pada skenario 1, peningkatan *max depth* dari 5 hingga 10 menghasilkan penurunan MAPE yang signifikan dari 2.922% menjadi 2.812%, dengan selisih sebesar 0.110%. Namun, saat *max depth* ditingkatkan lagi ke 20, MAPE hanya sedikit menurun menjadi 2.808%, dengan perbedaan sebesar 0.004%, menunjukkan bahwa peningkatan lebih lanjut dalam kedalaman tidak terlalu memengaruhi penurunan MAPE. Sebaliknya, skenario 2 menunjukkan MAPE terendah sebesar 2.836% pada *max depth* 10, tetapi kembali naik menjadi 2.847% pada *max depth* 20. Ini mengindikasikan bahwa performa skenario 2 menjadi kurang optimal pada kedalaman yang lebih tinggi. Dari segi standar deviasi, skenario 1 menunjukkan prediksi yang lebih stabil dengan standar deviasi

lebih rendah pada *max depth* 5 (0.052949) dan 10 (0.066677), menunjukkan konsistensi prediksi yang lebih baik. Pada *max depth* 20, standar deviasi skenario 1 meningkat menjadi 0.067006, yang menandakan adanya sedikit peningkatan variasi dalam hasil prediksi. Di sisi lain, skenario 2 memiliki standar deviasi yang lebih tinggi secara keseluruhan, terutama pada *max depth* 20 dengan nilai 0.071654, yang menunjukkan bahwa prediksi dalam skenario ini menjadi lebih tidak konsisten pada kedalaman yang lebih tinggi. Secara keseluruhan, perbedaan rata-rata MAPE antara kedua skenario cukup signifikan, Selisih rata-rata MAPE antara skenario 1 dan skenario 2 sebesar 0.030% (2.847% pada skenario 1 dan 2.877% pada skenario 2). Ini menunjukkan bahwa agregat mean pada skenario 1 lebih efektif dalam meningkatkan akurasi dengan kedalaman yang lebih besar, sementara skenario 2 cenderung mengalami variasi yang lebih besar, terutama pada kedalaman maksimum yang lebih tinggi.

Bootstrap samples menentukan proporsi data yang digunakan untuk membuat setiap *trees* dalam *forest*. Berikut pengaruhnya,

Tabel 4.6 Rata-rata MAPE parameter *Bootstrap Samples*

| | Parameter <i>Bootstrap Sample</i> | Rata-rata MAPE | Standar Deviasi MAPE |
|-------------------|--|-----------------------|-----------------------------|
| Skenario 1 | 0.25 | 2.849% | 0.075241 |
| | 0.5 | 2.845% | 0.082866 |
| | 0.7 | 2.847% | 0.088206 |
| Skenario 2 | 0.25 | 2.873% | 0.085276 |
| | 0.5 | 2.879% | 0.082941 |
| | 0.7 | 2.889% | 0.088051 |



Gambar 4.10 Visualisasi rata-rata MAPE pada parameter *Bootstrap Samples*

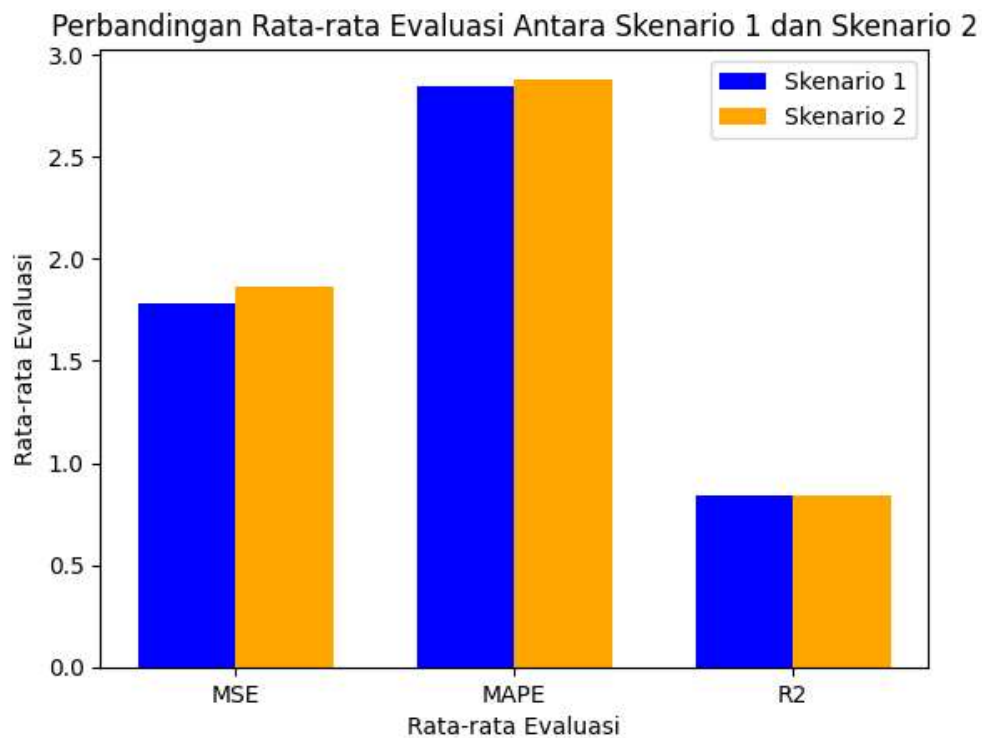
Pada gambar 4.10 dan tabel 4.6 menunjukkan bahwa pada setiap nilai *bootstrap samples* (0.25, 0.5, 0.7), skenario 1 secara konsisten menghasilkan MAPE yang lebih rendah dibandingkan skenario 2. Pada skenario 1, MAPE mengalami sedikit penurunan ketika *bootstrap samples* meningkat dari 0.25 (2.849%) ke 0.5 (2.845%), dengan selisih 0.004%. Namun, pada *bootstrap samples* 0.7, MAPE sedikit naik menjadi 2.847% hanya selisih 0.002%, menunjukkan bahwa peningkatan nilai *bootstrap* di atas 0.5 tidak memberikan keuntungan tambahan signifikan dalam pengurangan kesalahan prediksi. Di sisi lain, skenario 2 memiliki MAPE yang lebih tinggi secara konsisten, dengan MAPE tertinggi sebesar 2.889% pada *bootstrap samples* 0.7. Dari segi standar deviasi, skenario 1 memiliki nilai standar deviasi terendah pada *bootstrap samples* 0.25 sebesar

0.075241, yang meningkat saat *bootstrap samples* dinaikkan, menandakan bahwa variasi prediksi cenderung meningkat seiring kenaikan *bootstrap samples*. Skenario 2 juga menunjukkan standar deviasi yang lebih tinggi secara keseluruhan, terutama pada *bootstrap samples* 0.7, dengan nilai 0.088051, yang menunjukkan prediksi pada skenario ini menjadi lebih bervariasi dan kurang stabil saat nilai *bootstrap samples* meningkat. Dengan ini, selisih rata-rata MAPE antara kedua skenario adalah 0.026% (2.847% pada skenario 1 dan 2.873% pada skenario 2), dengan skenario 1 lebih unggul dalam hal akurasi pada semua nilai *bootstrap samples*. Selain itu, skenario 1 menunjukkan kestabilan yang lebih baik, terutama pada *bootstrap samples* yang lebih rendah, sedangkan skenario 2 menunjukkan variasi yang lebih tinggi pada *bootstrap samples* yang lebih besar. Hal ini mengindikasikan bahwa agregat *mean* pada skenario 1 lebih efektif dalam menangani variasi dalam *bootstrap sampling* dibandingkan agregat median.

Secara keseluruhan, skenario 1 dengan agregat *mean* terbukti lebih unggul dibandingkan skenario 2 dalam hal akurasi dan kestabilan. Pada berbagai pengaturan parameter seperti *train ratio*, *trees*, *max depth*, dan *bootstrap samples*, skenario 1 konsisten menghasilkan MAPE yang lebih rendah dan standar deviasi yang lebih stabil, menunjukkan akurasi dan konsistensi prediksi yang lebih baik. Sementara skenario 2 memiliki kelebihan dalam beberapa kasus, agregat *mean* di skenario 1 lebih efektif secara keseluruhan dalam menjaga performa model *Random Forest* pada data dan parameter yang dianalisis.

Tabel 4.7 Rata-rata evaluasi skenario 1 dan skenario 2

| Skenario | Rata-rata MSE | Rata-rata MAPE | Rata-rata R2 |
|----------|---------------|----------------|--------------|
| 1 | 1.779795 | 2.847% | 0.844061 |
| 2 | 1.860393 | 2.88% | 0.837018 |



Gambar 4.11 Visualisasi perbandingan rata-rata evaluasi skenario 1 dan skenario 2

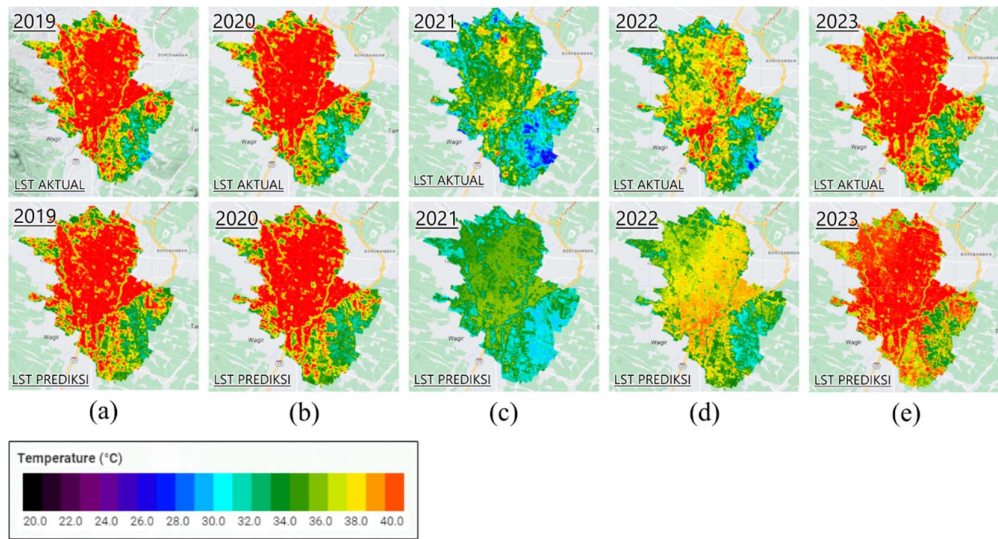
Gambar 4.11 dan tabel 4.7 menunjukkan seluruh rata-rata evaluasi per-skenario. Pada kedua skenario, hasil menunjukkan bahwa agregasi *mean* cenderung memberikan MSE yang lebih rendah dibandingkan agregasi *median*, menunjukkan bahwa prediksi model dengan *mean* lebih akurat dalam sebagian besar konfigurasi

parameter. Dalam konteks penelitian ini, model dengan agregasi *mean* juga menunjukkan performa lebih unggul pada nilai MAPE, yang merefleksikan kesalahan relatif. Meskipun nilai MAPE di kedua skenario relatif rendah (mengindikasikan prediksi yang cukup akurat), hasilnya sedikit lebih baik di skenario 1. Hal ini menandakan bahwa agregasi *mean* lebih efektif dalam menangani data dan menghasilkan prediksi akurat untuk studi UHI ini. Hal ini juga sama pada metrik R^2 yang rata-rata pada skenario 1 menghasilkan 0.844061 dan skenario 2 dengan nilai rata-rata 0.837018 yang menunjukkan hasil yang lebih tinggi pada skenario 1 yang berarti skenario 1 dengan baik menjelaskan variasi data. Dengan data pada grafik tersebut dapat disimpulkan bahwa skenario 1 menghasilkan nilai yang lebih baik pada seluruh metrik di penelitian ini.

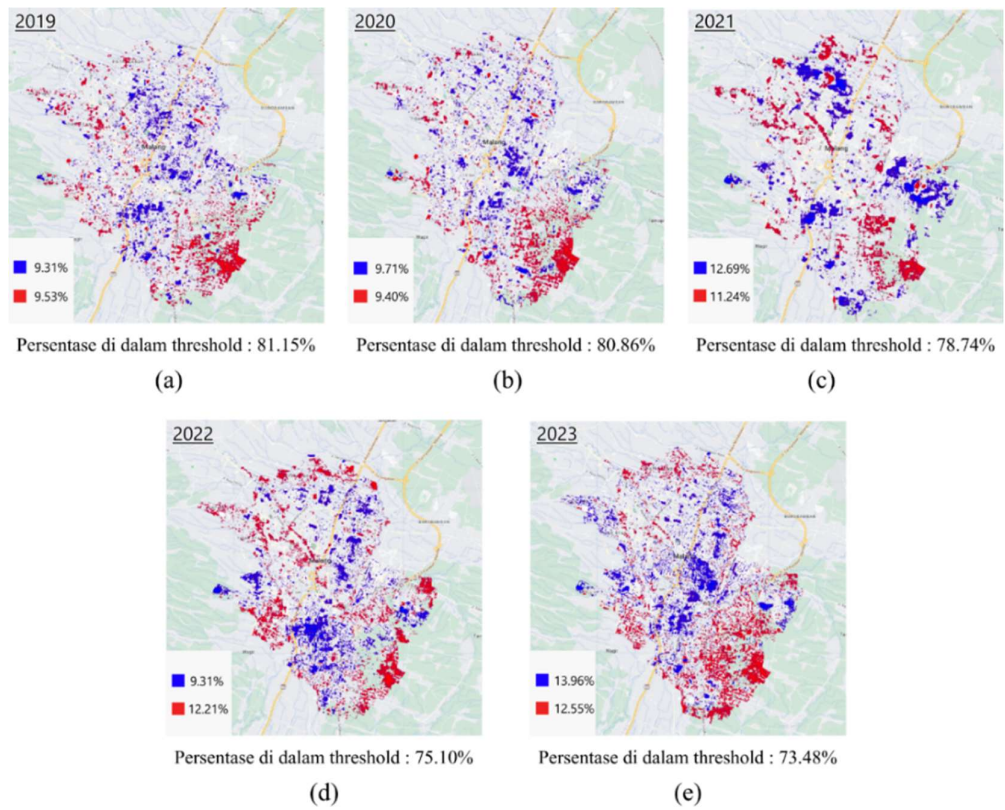
Berdasarkan hasil analisa evaluasi, model terbaik yang didapat dari keseluruhan skenario digabung adalah **Model ke-107 dalam skenario 1** dengan parameter rasio data 90:10, jumlah 500 *trees*, *max depth* 20, dan *bootstrap sample* sebesar 0.5. Model ini memiliki MSE terendah (1.56068), MAPE (2.700732), dan R^2 tertinggi (0.857488), menandakan akurasi prediksi yang tinggi dipilih sebagai model terbaik karena memberikan hasil paling optimal dibandingkan model lainnya.

4.3 Perbandingan Hasil Visualisasi

Selanjutnya, model terbaik yang telah dipilih akan dibandingkan dengan data aktual melalui visualisasi citra satelit, memperlihatkan prediksi UHI untuk setiap tahunnya.



Gambar 4.12 Visualisasi citra perbandingan LST aktual dan prediksi 2019-2023



LEGENDA : ■ Error dibawah threshold -2° ■ Error diatas threshold 2°

Gambar 4.13 Analisa persebaran *high error* prediksi UHI 2019-2023

Gambar 4.13 Menunjukkan persebaran area dengan *high error* dalam prediksi UHI di Kota Malang dari tahun 2019 hingga 2023. Pada tahun 2019 (a), area dengan prediksi akurat dalam *threshold* -2°C hingga 2°C mencapai 81.15%, menandakan performa model yang sangat baik pada tahun tersebut. Dari tahun ke tahun, area dengan prediksi akurat ini menurun secara bertahap, menjadi 80.86% pada 2020 (b), 78.74% pada 2021 (c), 75.10% pada 2022 (d), dan 73.48% pada 2023 (e). Penurunan ini mungkin mengindikasikan bahwa karakteristik data yang dipengaruhi oleh perkembangan lingkungan dan urbanisasi mulai memberikan kesulitan adaptasi model dikarenakan perubahan variasi data pada pengujian pertahunnya. Seperti yang dapat dilihat pada Gambar 4.12 yang membandingkan data aktual dan prediksi pertahunnya, kita dapat mengetahui terjadi perubahan variasi data pertahun terutama pada tahun 2021-2023 yang mengalami perubahan suhu pada data aktual secara signifikan dikarenakan pada saat itu terjadinya wabah virus COVID-19 sehingga aktivitas manusia menjadi berkurang.

Pada aspek perbandingan area di atas dan di bawah *threshold*, terlihat bahwa area dengan prediksi di atas *threshold* (lebih dari 2°C) mengalami peningkatan, terutama dari tahun 2021 hingga 2023, misalnya dari 11.24% pada 2021 menjadi 12.55% pada 2023. Ini bisa terkait dengan peningkatan aktivitas urban atau pola suhu yang mengalami kenaikan di beberapa area tertentu. Di sisi lain, area di bawah *threshold* (kurang dari -2°C) juga meningkat hingga 13.96% pada 2023, yang mungkin mengindikasikan adanya area dengan suhu yang lebih rendah dari perkiraan model, seperti area dengan lebih banyak vegetasi atau perubahan penggunaan lahan menjadi ruang hijau. Secara keseluruhan, perubahan ini

mencerminkan kebutuhan untuk menyesuaikan parameter atau data terbaru agar model dapat lebih efektif dalam memprediksi pola suhu yang dinamis di Kota Malang.

4.4 Integrasi dalam Islam

Integrasi antara sains dan Islam bukanlah hanya sekadar mengutip dalil-dalil islam dalam konteks ilmiah, atau bertujuan untuk mengislamkan sains itu sendiri. Tujuan utamanya adalah membuat para penggunanya lebih taat beragama melalui ilmu pengetahuan. Dengan demikian, kemajuan teknologi dan sains dapat digunakan sebagai alat untuk membantu manusia memenuhi tujuan penciptaannya, yaitu sebagai hamba Allah yang senantiasa beribadah dan berserah diri kepada-Nya, serta sebagai khalifah di muka bumi yang bertanggung jawab dalam memanfaatkan, menjaga, dan mengambil pelajaran dari alam. Untuk melaksanakan tanggung jawab ini, manusia perlu berinteraksi dengan Allah, sesama manusia, dan alam melalui prinsip-prinsip Muamalah Ma'a Allah, Muamalah Ma'a an-Nas, dan Muamalah Ma'a Alam.

4.4.1 Muamalah Ma'a Allah

Muamalah Ma'a Allah merujuk pada hubungan dan interaksi manusia dengan Allah SWT. Dalam konteks ini, interaksi mencakup ketaatan dalam menjalankan perintah-Nya, menjauhi larangan-Nya, dan berusaha mendekatkan diri kepada-Nya melalui ibadah dan amal saleh. Hal ini mencerminkan kesadaran seseorang terhadap keberadaan dan keesaan Allah SWT serta komitmennya untuk

menjalankan kehidupan sesuai dengan kehendak-Nya. Pada surat Al-Baqarah ayat 21 yang berbunyi,

يَا أَيُّهَا النَّاسُ اعْبُدُوا رَبَّكُمُ الَّذِي خَلَقَكُمْ وَالَّذِينَ مِنْ قَبْلِكُمْ لَعَلَّكُمْ تَتَّقُونَ ٢١

“Wahai manusia, sembahlah Tuhanmu yang telah menciptakan kamu dan orang-orang yang sebelum kamu agar kamu bertakwa.” (QS. Al-Baqarah: 21)

Menurut tafsir Ibnu Katsir, Allah SWT menjelaskan tentang sifat keesaan-Nya (uluhiyyah), yaitu bahwa Dia adalah satu-satu-Nya yang memberikan nikmat kepada hamba-hamba-Nya dengan menciptakan mereka dari ketiadaan ke dalam kehidupan, lalu melimpahkan kepada mereka berbagai macam nikmat, baik yang lahir maupun yang batin. Ayat ini ditujukan kepada semua golongan, baik orang-orang kafir maupun orang-orang munafik. Artinya, esakanlah Tuhan kalian yang telah menciptakan kalian dan orang-orang sebelum kalian. Hal serupa juga dijelaskan dalam riwayat lain dari Ibnu Abbas mengenai makna firman Allah pada surah Al-Baqarah ayat 22,

الَّذِي جَعَلَ لَكُمْ الْأَرْضَ فِرَاشًا وَالسَّمَاءَ بِنَاءً وَأَنْزَلَ مِنَ السَّمَاءِ مَاءً فَأَخْرَجَ بِهِ مِنَ الثَّمَرَاتِ رِزْقًا لَكُمْ فَلَا تَجْعَلُوا لِلَّهِ أَنْدَادًا وَأَنْتُمْ تَعْلَمُونَ ٢٢

“(Dialah) yang menjadikan bagimu bumi (sebagai) hamparan dan langit sebagai atap, dan Dialah yang menurunkan air (hujan) dari langit, lalu Dia menghasilkan dengan (hujan) itu buah-buahan sebagai rezeki untuk kamu. Oleh karena itu, janganlah kamu mengadakan sekutu-sekutu bagi Allah, padahal kamu mengetahui.” (QS. Al-Baqarah: 22)

Dijelaskan bahwa, janganlah kalian menyekutukan Allah dengan apa pun yang tidak dapat mendatangkan mudarat atau manfaat, padahal kalian mengetahui bahwa tidak ada Tuhan selain Allah yang memberikan rezeki kepada kalian. Oleh karena itu kita harus senantiasa mentaati perintah beribadah agar bertakwa dan dapat

memelihara diri serta terhindar dari murka dan siksa Allah. Dengan beribadah, berarti kita telah mempersiapkan diri untuk mengagungkan Allah, sehingga jiwa menjadi suci dan tunduk kepada kebenaran.

Dalam konteks muamalah ma'a Allah dan penelitian mengenai *Urban Heat Island*, konsep beribadah ini bisa dilakukan sebagai tanggung jawab manusia dalam menjaga bumi yang merupakan ciptaan Allah SWT. Fenomena UHI merupakan bentuk kerusakan lingkungan akibat perilaku manusia yang tidak sesuai dengan prinsip menjaga kelestarian alam. Penelitian UHI ini, yang berfokus pada pengurangan dampak negatif peningkatan suhu di perkotaan, sejalan dengan perintah untuk bertakwa dan memelihara diri.

Dengan melakukan penelitian ini, kita dapat memahami lebih dalam penyebab dan solusi untuk mengurangi fenomena UHI, yang berarti kita ikut berperan dalam menjaga kelestarian bumi sebagai bentuk ibadah kepada Allah SWT. Memahami dan menjaga keseimbangan alam adalah salah satu cara manusia menunjukkan kepatuhan kepada Allah SWT dan menghindari perbuatan yang dapat mendatangkan murkanya. Oleh karena itu, penelitian ini tidak hanya merupakan upaya ilmiah tetapi juga bentuk nyata dari pelaksanaan perintah Allah SWT untuk menjaga dan melestarikan ciptaannya, yang sejalan dengan tujuan beribadah untuk mencapai ketakwaan dan menjaga diri dari siksa-Nya.

4.4.2 Muamalah Ma'an-Nas

Muamalah Ma'an-Nas merujuk pada interaksi dan hubungan antar sesama manusia. Interaksi ini mencakup sikap saling menghormati, tolong-menolong, berlaku adil, dan berbuat baik. Prinsip ini mengedepankan pentingnya menjalani

kehidupan sosial yang harmonis, penuh kasih sayang, dan adil, sesuai dengan ajaran Islam maka kita harus berbuat baik kepada sesama manusia. Telah dijelaskan pada surat Al-Isra ayat 7 yang berbunyi,

إِنْ أَحْسَنْتُمْ أَحْسَنْتُمْ لِأَنْفُسِكُمْ وَإِنْ أَسَأْتُمْ فَلَهَا فَإِذَا جَاءَ وَعْدُ الْآخِرَةِ لِيَسُوءُوا وُجُوهَكُمْ وَلِيَدْخُلُوا الْمَسْجِدَ
كَمَا دَخَلُوهُ أَوَّلَ مَرَّةٍ وَلِيُتَبِّرُوا مَا عَلَوْا تَتْبِيرًا ۗ

“Jika berbuat baik, (berarti) kamu telah berbuat baik untuk dirimu sendiri. Jika kamu berbuat jahat, (kerugian dari kejahatan) itu kembali kepada dirimu sendiri. Apabila datang saat (kerusakan) yang kedua, (Kami bangkitkan musuhmu) untuk menyuramkan wajahmu, untuk memasuki masjid (Baitulmaqdis) sebagaimana memasukinya ketika pertama kali, dan untuk membinasakan apa saja yang mereka kuasai. (QS. Al-Isra: 7)

Tafsir oleh Ibnur Katsir menjelaskan bahwa setiap kebaikan yang kita lakukan maka akan menimpa kita sendiri dan jika kita melakukan kejahatan maka kejahatan itu juga akan menimpa diri kita sendiri. Oleh karena itu kita harus senantiasa berbuat baik kepada sesama manusia dan tidak melakukan kejahatan karena semua yang kita lakukan Allah SWT akan mengembalikannya ke diri kita sendiri. Bermanfaat bagi orang lain merupakan salah satu tindakan berbuat baik kepada sesama manusia. Pada sebuah hadits yang di riwayatkan Imam Ahmad, Rasulullah SAW bersabda,

خَيْرُ النَّاسِ أَنْفَعُهُمْ لِلنَّاسِ

“Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia (lainnya)”(HR.Ahmad).

hadist ini menegaskan bahwa kualitas seseorang tidak hanya diukur dari ibadahnya kepada Allah SWT, tetapi juga dari kontribusinya terhadap kebaikan masyarakat dan lingkungan sekitarnya. Hadist ini mendorong setiap individu untuk

aktif berpartisipasi dalam upaya saling membantu, memberikan manfaat, dan meringankan beban orang lain, baik melalui ilmu, tenaga, harta, maupun nasihat yang baik. Dalam konteks ini, segala bentuk kontribusi yang positif, seperti membantu orang yang membutuhkan, menjaga lingkungan, atau berinovasi demi kesejahteraan bersama, adalah bentuk nyata dari pelaksanaan muamalah ma'a anas.

Pada penelitian ini sejalan dengan penjelasan diatas karena bertujuan untuk memberikan manfaat yang luas bagi masyarakat, khususnya dalam konteks mitigasi fenomena *Urban Heat Island* (UHI) di Kota Malang. Dengan menggunakan metode prediksi yang lebih akurat, penelitian ini tidak hanya menawarkan solusi untuk memahami dan mengurangi dampak negatif dari UHI, tetapi juga berkontribusi pada peningkatan kualitas hidup masyarakat. Upaya untuk memahami dan mengatasi permasalahan UHI melalui pendekatan ilmiah menunjukkan komitmen untuk menjaga kelestarian lingkungan dan mengurangi risiko kesehatan yang diakibatkan oleh peningkatan suhu di perkotaan.

4.4.3 Muamalah Ma'a Alam

Muamalah Ma'a Alam merujuk pada interaksi dan hubungan manusia dengan lingkungan alam sekitarnya. Prinsip ini mencakup seperti tanggung jawab menjaga, memelihara, dan tidak merusak alam. Islam mengajarkan bahwa manusia memiliki tanggung jawab sebagai khalifah di bumi untuk menjaga kelestarian dan keseimbangan alam, serta mengelola sumber daya alam dengan bijaksana. Seperti pada penelitian ini yaitu fenomena *Urban Heat Island* merupakan fenomena alam yang terjadi karena ulah manusia yang membangun bangunan pada lahan kota

dengan jumlah yang begitu besar sehingga mengakibatkan keadaan alam menjadi tidak seimbang sehingga membuat suhu meningkat pada suatu wilayah. Sebagai khalifah penjaga kelestarian bumi manusia harus menjaga keadaan alam yang telah dititipkan Allah SWT kepada kita. Dalam surat Al-a-raf ayat 56 yang berbunyi,

وَلَا تُفْسِدُوا فِي الْأَرْضِ بَعْدَ إِصْلَاحِهَا وَادْعُوهُ حَوْفًا وَطَمَعًا إِنَّ رَحْمَتَ اللَّهِ قَرِيبٌ مِّنَ الْمُحْسِنِينَ ٥٦

"Janganlah kamu berbuat kerusakan di bumi setelah diatur dengan baik. Berdoalah kepada-Nya dengan rasa takut dan penuh harap. Sesungguhnya rahmat Allah sangat dekat dengan orang-orang yang berbuat baik." (QS. Al-A'raf: 56)

Menurut tafsir oleh Ibnu Katsir pada kitabnya Tafsir Al-Quranul Adhim dijelaskan bahwa Allah SWT melarang melakukan perbuatan yang dapat merusak bumi dan hal-hal yang membawa kerugian setelah terciptanya perbaikan. Jika suatu keadaan sudah berjalan sesuai dengan kelestariannya dan kemudian terjadi kerusakan, hal itu akan berdampak buruk bagi semua makhluk. Oleh karena itu, Allah SWT melarang tindakan semacam itu dan memerintahkan hamba-Nya untuk menyembah, berdoa, merendahkan diri, dan tunduk kepadanya. Allah SWT juga berfirman: (berdoalah kepadanya dengan rasa takut dan penuh harap. Sesungguhnya rahmat Allah sangat dekat dengan orang-orang yang berbuat baik) yakni, berdoalah dengan perasaan takut akan siksa yang mungkin ada di sisi-Nya dan harapan akan pahala berlimpah yang disediakan-Nya.

Penjelasan tafsir Ibnu Katsir di atas sangat relevan dalam konteks fenomena *Urban Heat Island*. Fenomena UHI sering kali disebabkan oleh aktivitas manusia yang tidak memperhatikan keseimbangan lingkungan. Manusia memiliki tanggung jawab untuk menjaga kelestarian bumi, termasuk menghindari tindakan yang dapat memperburuk kondisi lingkungan seperti UHI. Oleh karena itu, penting bagi

manusia untuk memelihara bumi sesuai dengan amanah tanggung jawab yang diberikan oleh Allah SWT dan selalu berdoa kepadanya.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian ini, penulis menyimpulkan bahwa pelatihan seluruh model untuk prediksi Urban Heat Island (UHI) di Kota Malang menggunakan metode Random Forest Regression menghasilkan dua skenario evaluasi, yaitu dengan agregasi mean dan median. Secara umum, model dengan agregasi mean menunjukkan performa yang lebih baik dalam hal akurasi prediksi, dengan nilai rata-rata MSE sebesar 1.779795 dan MAPE 2.847%, lebih rendah dibandingkan rata-rata MSE sebesar 1.860393 dan MAPE 2.88% pada skenario agregasi *median*. Perbedaan nilai MSE dan MAPE antara kedua skenario ini masing-masing sebesar 0.0806 dan 0.033%, yang menunjukkan bahwa model dengan agregasi *mean* menghasilkan prediksi yang lebih akurat dan konsisten, serta menjadi pilihan optimal berdasarkan hasil evaluasi metrik.

Sehingga model terbaik dalam penelitian ini dicapai pada konfigurasi **model ke-107, dengan rasio data latih 90%, jumlah tree 500, max depth 20, dan bootstrap samples 0.5**. Model ini mencatat nilai MSE terendah sebesar 1.56068, MAPE sebesar 2.700732, dan R^2 tertinggi sebesar 0.857488, yang menunjukkan kemampuannya dalam menjelaskan variabilitas data UHI di Kota Malang secara optimal. Agregasi *mean* juga menunjukkan stabilitas yang lebih baik dalam prediksi, yang terlihat dari standar deviasi yang lebih rendah pada berbagai konfigurasi parameter. Dengan stabilitas ini, model agregasi *mean* lebih *robust*

terhadap variasi data dan parameter, menjadikannya pilihan yang lebih sesuai untuk data UHI di Kota Malang.

Selain itu, penerapan model terbaik ini dalam memprediksi persebaran UHI per tahun (2019-2023) menunjukkan adanya penurunan bertahap dalam area dengan prediksi akurat pada rentang -2°C hingga 2°C , dari 81.15% pada 2019 menjadi 73.48% pada 2023. Penurunan ini mengindikasikan bahwa perubahan kondisi lingkungan dan urbanisasi secara bertahap mempengaruhi model dalam menghadapi variasi pola suhu. Area dengan prediksi di atas 2°C cenderung meningkat, yang mungkin disebabkan oleh aktivitas urbanisasi, sedangkan peningkatan area di bawah -2°C dapat dikaitkan dengan perkembangan vegetasi atau ruang hijau.

Meskipun agregasi *median* masih relevan pada situasi tertentu yang membutuhkan kekebalan terhadap *outliers*, dalam penelitian ini, agregasi *mean* dipilih sebagai metode terbaik. Temuan ini menekankan pentingnya pemilihan metode agregasi yang sesuai dengan karakteristik data untuk mencapai prediksi optimal serta mempertimbangkan dinamika perubahan lingkungan dalam memprediksi UHI.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, penulis memberikan beberapa saran yang disampaikan untuk penelitian selanjutnya, Berikut ini saran-saran yang diberikan:

1. Pada penelitian selanjutnya, dianjurkan untuk menambahkan beberapa variasi tuning parameter dan agregasi untuk metode *Random Forest Regression* agar dapat menghasilkan model yang lebih akurat lagi.
2. Menggunakan metode lain sebagai perbandingan dan eksplorasi untuk menemukan hasil yang terbaik.
3. Menambahkan fitur atau variabel untuk data prediksi UHI.
4. Pada hasil penerapan visualiasi citra dapat ditambahkan penjelasan yang lebih detail seperti penjelasan per-titik daerah atau kecamatan.

DAFTAR PUSTAKA

- Adi Nugraha, A. S., & Atmaja, D. M. (2020). Pemanfaatan Citra Penginderaan Jauh Multi-Temporal Untuk Deteksi Urban Heat Island (Uhi) Terhadap Perubahan Penggunaan Lahan Di Kabupaten Buleleng. *Majalah Ilmiah Globe*, 22(2), 71. <https://doi.org/10.24895/mig.2020.22-2.1046>
- Alexander, C. (2020). Normalised difference spectral indices and urban land cover as indicators of land surface temperature (LST). *International Journal of Applied Earth Observation and Geoinformation*, 86(November 2019), 102013. <https://doi.org/10.1016/j.jag.2019.102013>
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 1–24. <https://doi.org/10.7717/PEERJ-CS.623>
- Dede, M., Widiawaty, M. A., Ramadhan, Y. R., Ismail, A., & Nurdian, W. (2021). Prediksi Suhu Permukaan Menggunakan Artificial Neural Network-Cellular Automata Di Wilayah Cirebon Dan Sekitarnya. *Seminar Nasional Geomatika, April*, 153. <https://doi.org/10.24895/sng.2020.0-0.1130>
- Dvorakova, K., Heiden, U., & Van Wesemael, B. (2021). Sentinel-2 exposed soil composite for soil organic carbon prediction. *Remote Sensing*, 13(9), 1–20. <https://doi.org/10.3390/rs13091791>
- Furuya, M. T. G., Furuya, D. E. G., de Oliveira, L. Y. D., da Silva, P. A., Cicerelli, R. E., Gonçalves, W. N., Junior, J. M., Osco, L. P., & Ramos, A. P. M. (2023). A machine learning approach for mapping surface urban heat island using environmental and socioeconomic variables: a case study in a medium-sized Brazilian city. *Environmental Earth Sciences*, 82(13). <https://doi.org/10.1007/s12665-023-11017-8>
- Gao, J. (2023). R-Squared (R²) – How much variation is explained? . *Research Methods in Medicine & Health Sciences*. <https://doi.org/10.1177/26320843231186398>
- Handayanto, R. T., Haryono, & Herlawati. (2023). Perubahan Kerapatan Vegetasi dan Penutup Lahan Terhadap Urban Heat Island (UHI) di Kota Bekasi. *Journal of Students' Research in Computer Science*, 4(1), 157–166. <https://doi.org/10.31599/jsrsc.v4i1.2655>
- Larasati, A. P., Rahman, B., & Kautsary, J. (2022). Pengaruh Perkembangan Perkotaan Terhadap Fenomena Pulau Panas (Urban Heat Island). *Jurnal Kajian Ruang*, 2(1), 35. <https://doi.org/10.30659/jkr.v2i1.20469>

- Malakhov, D., Malakhov, D. V., & Tsyhuyeva, N. Y. (2020). Calculation of the biophysical parameters of vegetation in an arid area of south-eastern Kazakhstan using the normalized difference moisture index (NDMI). *Central Asian Journal of Environmental Science and Technology Innovation*, 4(4), 189–198. <https://www.researchgate.net/publication/348162851>
- Martin-Vide, J., Sarricolea, P., & Moreno-García, M. C. (2015). On the definition of urban heat island intensity: The “rural” reference. *Frontiers in Earth Science*, 3(June), 2–4. <https://doi.org/10.3389/feart.2015.00024>
- Mirzaei, P. A., & Haghighat, F. (2010). Approaches to study Urban Heat Island - Abilities and limitations. *Building and Environment*, 45(10), 2192–2201. <https://doi.org/10.1016/j.buildenv.2010.04.001>
- Permata, F. D., Putra, Y. S., & Adriat, R. (2023). Distribusi Spasial Tingkat Kebasahan Lahan di Kota Pfdontianak Menggunakan Normalized Difference Water Index (NDWI). *Prisma Fisika*, 10(3), 425. <https://doi.org/10.26418/pf.v10i3.60549>
- Pranata, S., & Kurniadin, N. (2021). Identifikasi Perubahan Indeks Kerapatan Bangunan Pasca Likuifaksi Di Kota Palu. *Buletin Poltanesa*, 22(1), 27–32. <https://doi.org/10.51967/tanesa.v22i1.469>
- Putra, C. D., Ramadhani, A., & Fatimah, E. (2021). Increasing Urban Heat Island area in Jakarta and it’s relation to land use changes. *IOP Conference Series: Earth and Environmental Science*, 737(1), 0–6. <https://doi.org/10.1088/1755-1315/737/1/012002>
- Rani, M., Kumar, P., Pandey, P. C., Srivastava, P. K., Chaudhary, B. S., Tomar, V., & Mandal, V. P. (2018). Multi-temporal NDVI and surface temperature analysis for Urban Heat Island inbuilt surrounding of sub-humid region: A case study of two geographical regions. *Remote Sensing Applications: Society and Environment*, 10, 163–172. <https://doi.org/10.1016/j.rsase.2018.03.007>
- Ravanelli, R., Nascetti, A., Cirigliano, R. V., Di Rico, C., Leuzzi, G., Monti, P., & Crespi, M. (2018). Monitoring the impact of land cover change on surface urban heat island through Google Earth Engine: Proposal of a global methodology, first applications and problems. *Remote Sensing*, 10(9), 1–21. <https://doi.org/10.3390/rs10091488>
- RIZWAN, A. M., DENNIS, L. Y. C., & LIU, C. (2008). A review on the generation, determination and mitigation of Urban Heat Island. *Journal of Environmental Sciences*, 20(1), 120–128. [https://doi.org/10.1016/S1001-0742\(08\)60019-4](https://doi.org/10.1016/S1001-0742(08)60019-4)
- Rose, M., & Hassen, H. R. (2019). *A Survey of Random Forest Pruning Techniques*. 99–109. <https://doi.org/10.5121/csit.2019.91808>
- Roy, M. H., & Larocque, D. (2012). Robustness of random forests for regression. *Journal of Nonparametric Statistics*, 24(4), 993–1006. <https://doi.org/10.1080/10485252.2012.715161>

- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, 20(1), 3–29. <https://doi.org/10.1177/1536867X20909688>
- T, D., & VK, K. (2018). Effect of Climate Change on Rice. In *Agrotechnology* (Vol. 07, Issue 02). <https://doi.org/10.4172/2168-9881.1000181>
- Tarippanah, F., & Ranjbar, A. (2021). Quantitative analysis of spatial distribution of land surface temperature (LST) in relation Ecohydrological, terrain and socio-economic factors based on Landsat data in mountainous area. *Advances in Space Research*, 68(9), 3622–3640. <https://doi.org/10.1016/j.asr.2021.07.008>
- Yao, Z., Jiafu, L., & Zhuyun, W. (2021). Predicting Surface Urban Heat Island in Meihokou City, China: A Combination Method of Monte Carlo and Random Forest. *Chinese Geographical Science*, 31(4), 659–670. <https://doi.org/10.1007/s11769-021->
- Zhao, W., Wu, H., Yin, G., & Duan, S. B. (2019). Normalization of the temporal effect on the MODIS land surface temperature product using random forest regression. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152(January), 109–118. <https://doi.org/10.1016/j.isprsjprs.2019.04.008>