

**PREDIKSI JUMLAH PENGUNJUNG HARIAN MUSEUM MPU PURWA
MENGUNAKAN METODE *LIGHT GRADIENT BOOSTING
MACHINE (LIGHTGBM)***

SKRIPSI

**Oleh:
CITRA AYU PUTRI NINGRUM
NIM. 220605110092**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

**PREDIKSI JUMLAH PENGUNJUNG HARIAN MUSEUM MPU PURWA
MENGUNAKAN METODE *LIGHT GRADIENT BOOSTING
MACHINE (LIGHTGBM)***

SKRIPSI

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

Oleh :
CITRA AYU PUTRI NINGRUM
220605110092

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

HALAMAN PERSETUJUAN

PREDIKSI JUMLAH PENGUNJUNG HARIAN MUSEUM MPU PURWA MENGUNAKAN METODE *LIGHT GRADIENT* *BOOSTING MACHINE (LIGHTGBM)*

SKRIPSI

Oleh:
CITRA AYU PUTRI NINGRUM
NIM. 220605110092

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 05 Mei 2026

Pembimbing I,


Prof. Dr. Muhammad Faisal, M.T.
NIP. 19740510 200501 1 007

Pembimbing II,


Dr. Zainal Abidin, M.Kom
NIP. 19760613 200501 1 004

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




Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

PREDIKSI JUMLAH PENGUNJUNG HARIAN MUSEUM MPU PURWA MENGUNAKAN METODE *LIGHT GRADIENT* *BOOSTING MACHINE (LIGHTGBM)*

SKRIPSI

CITRA AYU PUTRI NINGRUM
NIM. 220605110092

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 11 Mei 2026

Susunan Dewan Penguji

Ketua Penguji : Fathurrohman, M.Kom
NIP. 19700731 200501 1 002

Anggota Penguji I : Johan Ericka Wahyu Prakasa, M.Kom
NIP. 19831213 201903 1 004

Anggota Penguji II : Prof. Dr. Muhammad Faisal, M.T.
NIP. 19740510 200501 1 007

Anggota Penguji III : Dr. Zainal Abidin, M.Kom
NIP. 19760613 200501 1 004

()

()

()

()

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



Agus Yono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Citra Ayu Putri Ningrum
NIM : 220605110092
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Prediksi Jumlah Pengunjung Harian Museum Mpu
Purwa Menggunakan Metode *Light Gradient
Boosting Machine (LightGBM)*

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, 05 Mei 2026

Yang membuat pernyataan,



Citra Ayu Putri Ningrum

NIM. 220605110092

MOTTO

... *“Even small steps are still progress”* ...

HALAMAN PERSEMBAHAN

Segala puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala
atas segala rahmat, karunia, serta petunjuk-Nya sehingga skripsi ini dapat
diselesaikan dengan baik.

Dengan penuh rasa hormat dan terima kasih, karya ini penulis persembahkan
kepada:

Kedua orang tua saya, ayah dan ibu tercinta
yang senantiasa memberikan doa, kasih sayang, dukungan, serta pengorbanan
yang tak ternilai sepanjang perjalanan pendidikan dan penyusunan skripsi ini.

Ketiga adik saya,
yang memberi saya motivasi besar untuk menyelesaikan skripsi ini

Seluruh keluarga besar,
yang senantiasa memberikan dukungan moral, doa, serta motivasi sehingga
penulis dapat menyelesaikan studi dan penyusunan skripsi ini dengan baik.

Dosen pembimbing dan dosen penguji,
yang telah meluangkan waktu, memberikan arahan, bimbingan, serta ilmu yang
sangat berharga selama proses penyusunan skripsi ini.

Teman-teman seperjuangan,
yang telah menemani perjalanan selama masa perkuliahan, berbagi pengalaman,
memberikan bantuan, dukungan, dan kebersamaan yang berkesan.

Semua pihak yang tidak dapat disebutkan satu per satu,
yang telah memberikan bantuan, doa, dan dukungan dalam berbagai bentuk
sehingga skripsi ini dapat terselesaikan.

Semoga segala kebaikan, dukungan, dan doa yang telah diberikan mendapatkan
balasan yang terbaik dari Allah Subhanahu Wa Ta'ala.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Alhamdulillah Rabbil 'Alamin, segala puji dan syukur penulis panjatkan ke hadirat Allah *Subhanahu Wa Ta'ala* atas segala rahmat, karunia, taufik, serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Prediksi Jumlah Pengunjung Harian Museum Mpu Purwa Menggunakan Metode *Light Gradient Boosting Machine*” dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada junjungan kita Nabi Muhammad *Shallallahu 'Alaihi Wasallam*, beserta keluarga, sahabat, dan seluruh umatnya hingga akhir zaman.

Penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, bimbingan, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., CAHRM., CRMP., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan kesempatan kepada penulis untuk menempuh pendidikan di lingkungan Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan dukungan dan fasilitas selama proses perkuliahan.
3. Supriyono M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan arahan dan dukungan selama masa studi.

4. Prof. Dr. Muhammad Faisal, M.T., selaku Dosen Pembimbing I yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan arahan, masukan, serta bimbingan yang sangat berharga selama proses penyusunan skripsi ini.
5. Dr. Zainal Abidin, M.Kom., selaku Dosen Pembimbing II yang telah memberikan bimbingan, saran, motivasi, dan berbagai masukan konstruktif sehingga skripsi ini dapat terselesaikan dengan baik.
6. Fatchurrohman, M.Kom. dan Johan Ericka Wahyu Prakasa, M.Kom., selaku dosen penguji yang telah memberikan kritik, saran, dan masukan yang sangat bermanfaat untuk penyempurnaan skripsi ini.
7. Nia Faricha, S.Si., selaku administrasi Program Studi Teknik Informatika yang senantiasa membantu dengan sabar dalam memberikan informasi, arahan, dan pelayanan akademik selama masa perkuliahan hingga penyelesaian skripsi.
8. *Support system* utama penulis yang penulis panggil ayah dan ibu, Anang Wahyudi dan Esti Kusuma Ningrum yang selalu memberikan kasih sayang, doa, dukungan, pengorbanan, serta motivasi yang tiada henti dalam setiap langkah perjalanan pendidikan penulis.
9. Ketiga adik penulis, Abdul Raffi Nur Wahyudi, Abrisam Ahza Nur Wahyudi, dan Fatima Syakila Putri Ningrum, yang merupakan salah satu alasan terbesar penulis untuk menyelesaikan skripsi ini.

10. Seluruh keluarga besar ayah dan ibu, yang senantiasa memberikan doa, dukungan, perhatian, dan motivasi kepada penulis selama menempuh pendidikan.
11. Sivana Firdausy Khurriyyah dan Fadhillah Putri Triana, sahabat terdekat penulis selama masa perkuliahan yang senantiasa menemani, mendukung, mendengarkan keluh kesah, serta menjadi tempat berbagi cerita dalam suka maupun duka selama menjalani kehidupan akademik dan keseharian di bangku perkuliahan
12. Nuril Nuzulia Afifah, Annisa Fatmawati Atu, dan Wahyu Windiastuti, teman-teman dekat penulis yang dipertemukan sejak awal memasuki dunia perkuliahan dan telah menjadi bagian dari berbagai momen berharga, baik dalam kegiatan perkuliahan maupun kebersamaan selama menjalani kehidupan di Kota Malang.
13. Atsila dan Silvi, yang telah menjadi teman berbagi cerita, berdiskusi, dan saling bertukar informasi terkait perkuliahan. Terima kasih atas dukungan, bantuan, serta kebersamaan yang diberikan selama menjalani masa studi.
14. Anggun dan Dafa, teman yang dipertemukan melalui kegiatan Kuliah Kerja Mahasiswa (KKM), yang telah memberikan banyak kenangan, kebersamaan, serta menjadi teman berbagi cerita dan pengalaman selama masa perkuliahan.
15. Sahabat-sahabat penulis yang tergabung dalam Keluarga PackMan, yaitu Fatih, Kamal, Tiok, Delfiana, Atira, dan Ussisa, yang senantiasa

memberikan dukungan, semangat, serta menjadi tempat berbagi cerita dan melepas penat selama perjalanan studi penulis.

16. Sahabat-sahabat penulis yang tergabung dalam Bocah⁷, yaitu Leny, Zihan, Elis, Maya, Alsa, dan Fira, yang selalu memberikan semangat, dukungan, doa, serta kebersamaan yang berarti selama proses perkuliahan hingga penyusunan skripsi ini.

17. Dan, kepada diri saya sendiri, Citra Ayu Putri Ningrum, yang telah menyelesaikan kembali satu bab dalam ceritanya. Terima kasih karena telah menyelesaikan bab ini dengan cukup baik, belajar untuk tidak memandang rendah diri sendiri, serta terus berusaha dan bertahan dalam setiap proses yang dilalui hingga sampai pada titik ini. Semoga langkah-langkah selanjutnya dapat dijalani dengan penuh keyakinan dan semangat.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari kata sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan dan pengembangan penelitian di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat serta kontribusi bagi pengembangan ilmu pengetahuan, khususnya di bidang Teknik Informatika.

Wassalamu'alaikum Wr, Wb.

Malang, 05 Mei 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvi
ABSTRAK	xvii
ABSTRACT	xviii
مستخلص البحث.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	7
1.3 Batasan Masalah.....	7
1.4 Tujuan Penelitian	8
1.5 Manfaat Penelitian	8
BAB II STUDI PUSTAKA	10
2.1 Penelitian Terdahulu	10
2.2 Time Series dan Forecasting	14
2.3 Metode LightGBM.....	15
2.4 Feature Engineering	19
2.5 Evaluasi Model.....	20
BAB III METODE PENELITIAN	22
3.1 Desain Penelitian.....	22
3.2 Desain Sistem.....	23
3.2.1 <i>Exploratory Data Analysis</i> (EDA)	24
3.2.2 <i>Preprocessing Data</i>	25
3.2.3 <i>Feature Engineering</i>	29
3.2.4 Modelling <i>LightGBM</i>	34
3.2.5 Skenario Pengujian.....	42
3.2.6 Evaluasi Model.....	44
BAB IV HASIL DAN PEMBAHASAN	47
4.1 Exploratory Data Analysis (EDA)	47
4.2 Hasil Preprocessing Data	54
4.2.1 Hasil <i>Cleaning Data</i>	54
4.2.2 Hasil <i>Agregation Data</i>	55
4.2.3 Hasil <i>Capping Data</i>	56
4.3 Feature Engineering	57
4.3.1 Hasil Ekstrasi Fitur Waktu	58
4.3.2 Hasil <i>Shock Feature</i>	59
4.3.3 Hasil <i>Lag Feature</i>	60
4.3.4 Hasil <i>Rolling Mean</i>	61

4.4 Skenario Pengujian.....	62
4.4.1 Skenario Pengujian A.....	62
4.4.2 Skenario Pengujian B.....	64
4.4.3 Skenario Pengujian C.....	67
4.5 Pembahasan.....	69
4.5.1 Analisis Perbandingan Skenario Pembagian Data	70
4.5.2 Analisis Hyperparameter.....	72
4.5.3 Analisis Metrik Evaluasi	74
4.5.4 Analisis Kesesuaian Metode <i>LightGBM</i>	75
4.6 Intergrasi Islam.....	76
BAB V KESIMPULAN DAN SARAN	82
5.1 Kesimpulan	82
5.2 Saran.....	82
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 2.1 Perbedaan Level-wise tree dan Leaf-wise tree.....	16
Gambar 3.1 Desain Penelitian.....	22
Gambar 3.2 Desain Sistem.....	23
Gambar 3.3 Alur Kerja Preprocessing Data.....	25
Gambar 3.4 Alur Kerja LightGBM.....	35
Gambar 3.5 Ilustrasi Pohon Keputusan Pertama.....	42
Gambar 4.1 Grafik Rata-Rata Jumlah Pengunjung.....	47
Gambar 4.2 Distribusi Jumlah Pengunjung	49
Gambar 4.3 Deteksi Outlier Jumlah Pengunjung.....	51
Gambar 4.4 Presentase Missing Value per Kolom	52
Gambar 4.5 Grafik Perbandingan Data Aktual dan Prediksi Skenario A	63
Gambar 4.6 Grafik Perbandingan Data Aktual dan Prediksi Skenario B	65
Gambar 4.7 Grafik Perbandingan Data Aktual dan Prediksi Skenario C	68
Gambar 4.8 Grafik Perbandingan Evaluasi Model	71

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	10
Tabel 3.1 Sebelum Cleaning Data.....	26
Tabel 3.2 Setelah Cleaning Data.....	26
Tabel 3.3 Sebelum Agregration Data.....	27
Tabel 3.4 Setelah Agregration Data	28
Tabel 3.5 Setelah Proses Capping.....	29
Tabel 3.6 Tabel data setelah ekstrasi fitur waktu	30
Tabel 3.7 Tabel Data setelah Lag Feature.....	33
Tabel 3.8 Sampel Data untuk simulasi perhitungan model LightGBM.....	36
Tabel 3.9 Tabel Data gradien	38
Tabel 3.10 Tabel Data Jumlah Prediksi terbaru	40
Tabel 3.11 Tabel Skenario Pengujian	43
Tabel 3.12 Tabel Hyperparameter.....	44
Tabel 4.1 Data setelah cleaning	54
Tabel 4.2 Data setelah Agregration.....	56
Tabel 4.3 Data Setelah Capping.....	57
Tabel 4.4 Hasil Ektrasi Fitur Waktu	58
Tabel 4.5 Hasil Shock Feature	59
Tabel 4.6 Hasil Lag Feature	60
Tabel 4.7 Hasil Rolling Mean	61
Tabel 4.8 Perbandingan Hyperparameter.....	72

ABSTRAK

Ningrum, Citra Ayu Putri. 2026. **Prediksi Jumlah Pengunjung Harian Museum Mpu Purwa Menggunakan Metode *Light Gradient Boosting Machine (LightGBM)***. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Muhammad Faisal, M.T. (II) Dr. Zainal Abidin, M.Kom.

Kata kunci: *Prediksi, Time Series, LightGBM, Feature Engineering, Museum Mpu Purwa.*

Prediksi jumlah pengunjung harian merupakan salah satu langkah penting dalam mendukung pengelolaan dan perencanaan operasional destinasi wisata. Penelitian ini bertujuan untuk memprediksi jumlah pengunjung harian Museum Mpu Purwa menggunakan metode *Light Gradient Boosting Machine (LightGBM)* berdasarkan data historis kunjungan periode 2022–2025. Tahapan penelitian meliputi *exploratory data analysis (EDA)*, *preprocessing data* yang terdiri dari *cleaning*, *aggregation*, dan *capping*, serta *feature engineering* berupa fitur kalender, *lag feature*, *rolling mean*, dan *shock feature* untuk menangkap pola temporal pada data deret waktu. Selanjutnya dilakukan proses pelatihan model dengan beberapa kombinasi *hyperparameter* dan tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. Kinerja model dievaluasi menggunakan metrik *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Hasil penelitian menunjukkan bahwa model *LightGBM* mampu menghasilkan prediksi yang baik terhadap jumlah pengunjung harian Museum Mpu Purwa, dengan skenario terbaik diperoleh pada pembagian data 90:10 yang menghasilkan nilai MAE sebesar 1,65, RMSE sebesar 2,27, dan MAPE sebesar 7,29%. Nilai tersebut menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang rendah dan mampu mengikuti pola fluktuasi jumlah pengunjung harian dengan baik. Dengan demikian, metode *LightGBM* dapat digunakan sebagai pendekatan yang efektif dalam memprediksi jumlah pengunjung harian serta mendukung pengambilan keputusan yang lebih tepat dalam pengelolaan Museum Mpu Purwa.

ABSTRACT

Ningrum, Citra Ayu Putri. 2026. **Daily Visitor Prediction of Museum Mpu Purwa Using the Light Gradient Boosting Machine (LightGBM) Method**. Undergraduate Thesis. Informatics Engineering Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Supervisors: (I) Prof. Dr. Muhammad Faisal, M.T. (II) Dr. Zainal Abidin, M.Kom.

Keywords: Prediction, Time Series, LightGBM, Feature Engineering, Museum Mpu Purwa.

Daily visitor prediction is an important step in supporting the management and operational planning of tourist destinations. This study aims to predict the daily number of visitors to Museum Mpu Purwa using the Light Gradient Boosting Machine (LightGBM) method based on historical visitor data from 2022 to 2025. The research stages include Exploratory Data Analysis (EDA), data preprocessing consisting of cleaning, aggregation, and capping, as well as feature engineering involving calendar features, lag features, rolling mean, and shock features to capture temporal patterns in time-series data. Subsequently, the model was trained using several hyperparameter combinations and three data-splitting scenarios, namely 70:30, 80:20, and 90:10. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results show that the LightGBM model is capable of generating accurate predictions for the daily number of visitors to Museum Mpu Purwa. The best performance was achieved using the 90:10 data-splitting scenario, resulting in an MAE of 1.65, an RMSE of 2.27, and a MAPE of 7.29%. These results indicate that the model has a low prediction error rate and is able to effectively capture fluctuations in daily visitor patterns. Therefore, the LightGBM method can be considered an effective approach for daily visitor prediction and can support more informed decision-making in the management of Museum Mpu Purwa.

مستخلص البحث

نينغروم، سيترا أيو بوتري. ٢٠٢٦. التنبؤ بعدد الزوار اليومي لمتحف مبو بوروا باستخدام خوارزمية تعزيز التدرج الخفيف للقرارات بحث تخرج. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية (LightGBM) الحكومية مالانج. المشرف الأول: الأستاذ الدكتور محمد فيصل، الماجستير؛ المشرف الثاني: الدكتور زين العابدين، الماجستير في علوم الحاسوب

الكلمات المفتاحية: التنبؤ، السلاسل الزمنية، تعزيز التدرج الخفيف للقرارات، هندسة الخصائص، متحف مبو بوروا

يُعدُّ التنبؤ بعدد الزوار اليومي من الخطوات المهمة في دعم إدارة الوجهات السياحية والتخطيط التشغيلي لها. وتهدف هذه اعتماداً (LightGBM) الدراسة إلى التنبؤ بعدد الزوار اليومي لمتحف مبو بوروا باستخدام خوارزمية تعزيز التدرج الخفيف للقرارات على البيانات التاريخية للزوار خلال الفترة من عام ٢٠٢٢ إلى عام ٢٠٢٥. وقد اشتملت مراحل البحث على تحليل البيانات الاستكشافي، والمعالجة المسبقة للبيانات التي تضمنت التنظيف والتجميع ومعالجة القيم المتطرفة، بالإضافة إلى هندسة الخصائص التي شملت خصائص التقويم، وخصائص التأخير، والمتوسط المتحرك، وخصائص الصدمات؛ وذلك لتمثيل الأنماط الزمنية في بيانات السلاسل الزمنية. بعد ذلك تم تدريب النموذج باستخدام عدة مجموعات من الملاحظات الفائقة وثلاثة سيناريوهات لتقسيم البيانات والجذر التربيعي (MAE) وهي ٧٠:٣٠ و ٨٠:٢٠ و ٩٠:١٠. وتم تقييم أداء النموذج باستخدام مقاييس متوسط الخطأ المطلق وأظهرت نتائج الدراسة أن نموذج تعزيز التدرج الخفيف (MAPE) ومتوسط نسبة الخطأ المطلق (RMSE) لمتوسط مربعات الخطأ للقرارات حقق أداءً جيداً في التنبؤ بعدد الزوار اليومي لمتحف مبو بوروا، حيث أظهر أفضل النتائج عند استخدام سيناريو تقسيم البيانات بنسبة ٩٠:١٠، محققاً قيمة متوسط خطأ مطلق بلغت ١,٦٥، وقيمة الجذر التربيعي لمتوسط مربعات الخطأ بلغت ٢,٢٧ وقيمة متوسط نسبة الخطأ المطلق بلغت ٧,٢٩٪. وتشير هذه النتائج إلى انخفاض معدل الخطأ في التنبؤ وقدرة النموذج على تمثيل التغيرات اليومية في أعداد الزوار بصورة جيدة. وبناءً على ذلك، يمكن استخدام هذه الخوارزمية بوصفها أسلوباً فعالاً للتنبؤ بأعداد الزوار اليومية، بما يساهم في دعم اتخاذ القرارات وتحسين إدارة متحف مبو بوروا

BAB I

PENDAHULUAN

1.1 Latar Belakang

Museum memiliki peran penting sebagai sarana pelestarian budaya sekaligus media edukasi sejarah bagi masyarakat. Selain menyimpan koleksi bersejarah, museum juga berfungsi sebagai sarana pembelajaran interaktif yang meningkatkan pemahaman masyarakat terhadap warisan budaya dan sejarah lokal (Mohamad et al., 2024). Fungsi edukatif ini semakin relevan karena museum kini bertransformasi secara virtual untuk menarik minat dan mempermudah akses informasi (Dahlan, 2024). Dengan pengelolaan koleksi yang tematik dan kronologis, museum mampu menciptakan pengalaman belajar yang berkesan serta berperan sebagai destinasi wisata edukatif (Zihad & Windari, 2025). Meningkatnya kunjungan memberikan dampak positif bagi pelestarian budaya dan pariwisata daerah. Museum Mpu Purwa, sebagai salah satu museum di Kota Malang, memiliki potensi besar, namun tetap menghadapi tantangan dalam mengelola jumlah pengunjung yang dinamis.

Namun demikian, jumlah pengunjung museum cenderung mengalami fluktuasi setiap harinya. Pada akhir pekan atau musim liburan, jumlah kunjungan meningkat tajam, sementara pada hari biasa cenderung mengalami penurunan (Julianto et al., 2021). Fluktuasi ini menimbulkan kendala dalam perencanaan fasilitas, penjadwalan kegiatan, dan distribusi sumber daya. Oleh karena itu, prediksi jumlah pengunjung menjadi penting untuk membantu pengelola museum dalam mengantisipasi perubahan pola kunjungan (Meimela, 2021)

Permasalahan fluktuasi jumlah pengunjung yang sulit diprediksi berpotensi menghambat optimalisasi fungsi museum sebagai pusat edukasi dan wisata budaya. Ketika terjadi lonjakan pengunjung tanpa adanya antipasi, fasilitas museum dapat mengalami keterbatasan, sementara pada saat penurunan kunjungan, potensi edukatif museum tidak dapat berjalan sebagaimana mestinya. Oleh karena itu, dibutuhkan suatu pendekatan ilmiah yang mampu memberikan prediksi mengenai jumlah pengunjung harian, sehingga pengelola dapat menyusun strategi yang lebih efektif dan berkelanjutan. Urgensi prediksi ini juga sejalan dengan prinsip Islam yang menekankan pentingnya perencanaan dan pengelolaan sumber daya secara bijak.

Al-Qur'an memberikan petunjuk mengenai pentingnya perencanaan dan pengambilan keputusan yang bijak melalui analisis terhadap tanda-tanda yang ada. Sebagaimana yang telah dijelaskan dalam Tafsir Al-Qur'an Kementerian Agama Republik Indonesia, kisah Nabi Yusuf a.s. mengandung pesan penting mengenai perencanaan dan pengambilan keputusan yang didasarkan pada analisis kondisi masa depan (Departemen Agama RI, 2008.) . Hal ini tergambar ketika Nabi Yusuf a.s. menafsirkan mimpi Raja Mesir sebagai bentuk antisipasi terhadap masa sulit yang akan dihadapi, sebagaimana firman Allah Swt. dalam QS. Yusuf ayat 47–49:

قَالَ تَزْرَعُونَ سَبْعَ سِنِينَ دَأَبًا فَمَا حَصَدْتُمْ فَذَرُوهُ فِي سُنْبُلِهِ ۖ إِلَّا قَلِيلًا مِّمَّا تَأْكُلُونَ ٤٧ ثُمَّ يَأْتِي مِنْ بَعْدِ ذَلِكَ سَبْعٌ شِدَادٌ يَأْكُلْنَ مَا قَدَّمْتُمْ لَهُنَّ إِلَّا قَلِيلًا مِّمَّا تُخْصِنُونَ ٤٨ ثُمَّ يَأْتِي مِنْ بَعْدِ ذَلِكَ عَامٌ فِيهِ يُغَاثُ النَّاسُ وَفِيهِ يَعْرِضُونَ ٤٩

٤٩

“Yusuf berkata: “Kalian akan bercocok tanam tujuh tahun sebagaimana biasa, maka apa yang kamu tuai hendaklah kamu biarkan di bulirnya kecuali sedikit untuk kamu makan. Kemudian setelah itu akan datang tujuh tahun yang sangat sulit, yang akan menghabiskan apa yang kamu simpan untuk menghadapinya, kecuali sedikit

dari yang kamu simpan. Kemudian setelah itu akan datang tahun yang padanya manusia diberi hujan dan pada masa itu mereka memeras anggur.” (Q.S. Yusuf [12]: 47–49)

Menurut Tafsir Ibnu Katsir, ayat ini menggambarkan bentuk perencanaan ekonomi dan strategi antisipatif berdasarkan informasi masa lalu untuk menghadapi kondisi masa depan. Nabi Yusuf a.s. tidak hanya menafsirkan mimpi, tetapi juga menyusun langkah-langkah konkret guna menjaga kestabilan pangan rakyat Mesir (Katsir, 2003.).

Sementara itu, dalam Tafsir Ath-Thabari, dijelaskan bahwa penafsiran Nabi Yusuf a.s. terhadap mimpi Raja Mesir merupakan bentuk pemahaman yang mendalam terhadap petunjuk Allah Swt. yang disampaikan melalui simbol-simbol mimpi. Ath-Thabari menekankan bahwa Nabi Yusuf a.s. tidak hanya menjelaskan makna mimpi tersebut, tetapi juga memberikan arahan konkret berupa langkah-langkah yang harus dilakukan, yaitu menyimpan sebagian hasil panen pada masa subur sebagai persiapan menghadapi masa paceklik (Abu Ja’far Muhammad bin Jarir ath-Thabari, 2007.). Penafsiran ini menunjukkan bahwa Islam mengajarkan pentingnya perencanaan, kehati-hatian, serta pengelolaan sumber daya secara terarah sebagai bentuk ikhtiar dalam menghadapi ketidakpastian di masa depan.

Adapun menurut Tafsir Al-Mishbah, Quraish Shihab menegaskan bahwa kisah Nabi Yusuf ini mengandung pesan moral dan strategis bagi umat manusia dalam menghadapi perubahan kondisi sosial dan ekonomi. Perencanaan yang dilakukan Nabi Yusuf bukan sekadar tindakan reaktif, tetapi bentuk upaya preventif yang didasarkan pada pengetahuan dan perhitungan matang. Quraish Shihab juga menyoroti bahwa keberhasilan Yusuf dalam mengelola masa subur dan paceklik

menjadi contoh nyata pentingnya *foresight* atau pandangan jauh ke depan dalam pengambilan keputusan (Shihab, 2005.).

Selain menekankan pentingnya perencanaan sebagaimana tergambar dalam kisah Nabi Yusuf a.s., Al-Qur'an juga memberikan gambaran tentang keteraturan dan ketepatan waktu dalam sistem ciptaan Allah. Fenomena alam seperti peredaran matahari dan bulan bukan hanya menunjukkan kekuasaan Allah, tetapi juga menjadi dasar manusia dalam melakukan penghitungan waktu dan perencanaan aktivitas kehidupan. Hal ini sebagaimana firman Allah dalam QS. Yasin ayat 39–40:

وَالْقَمَرَ قَدَرْنَاهُ مَنَازِلَ حَتَّىٰ عَادَ كَالْعُرْجُونِ الْقَدِيمِ ۚ لَا الشَّمْسُ يَنْبَغِي لَهَا أَنْ تُدْرِكَ الْقَمَرَ وَلَا اللَّيْلُ سَابِقُ النَّهَارِ ۚ وَكُلٌّ فِي فَلَكٍ يَسْبَحُونَ ٤٠

“Dan telah Kami tetapkan bagi bulan manzilah-manzilah (tempat-tempat peredarannya), sehingga (setelah sampai ke tempat peredaran yang terakhir) kembalilah dia seperti tandan yang tua. Tidaklah mungkin bagi matahari mendapatkan bulan dan malam pun tidak dapat mendahului siang. Dan masing-masing beredar pada garis edarnya.” (Q.S. Yasin [36]: 39–40).

Al-Qur'an juga menyinggung pentingnya sistem perhitungan waktu melalui keteraturan alam semesta, sebagaimana dijelaskan dalam QS. Yasin ayat 39, yang menggambarkan peredaran bulan sebagai tanda kebesaran Allah dan sarana manusia dalam mengatur waktu dan perencanaan hidupnya. Menurut Tafsir Ibnu Katsir ayat ini menunjukkan bahwa Allah telah menetapkan bagi bulan tahapan-tahapan peredarannya sehingga manusia dapat mengetahui hitungan waktu, bulan, dan tahun (Ibnu Katsir, 2004). Keteraturan ini menjadi bukti kebijaksanaan Ilahi sekaligus dorongan bagi manusia untuk berpikir sistematis, terencana, dan memperhitungkan setiap langkah berdasarkan tanda-tanda yang telah Allah

ciptakan. Prinsip keteraturan dan perhitungan inilah yang dapat dianalogikan dengan upaya manusia dalam mengembangkan sistem prediksi berbasis data untuk memahami pola di masa depan.

Dengan demikian, nilai-nilai perencanaan dan pengambilan keputusan berdasarkan analisis yang terkandung dalam kisah Nabi Yusuf a.s. serta keteraturan sistem alam sebagaimana dijelaskan dalam QS. Yasin menjadi dasar spiritual dan konseptual bagi penelitian ini.

Upaya melakukan prediksi jumlah pengunjung museum sejalan dengan kisah Nabi Yusuf a.s. dalam QS. Yusuf ayat 47–49, ketika beliau menafsirkan mimpi Raja Mesir dan menyusun strategi penyimpanan hasil panen sebagai langkah antisipatif menghadapi masa paceklik. Menurut Tafsir Ibnu Katsir, kisah tersebut menggambarkan pentingnya perencanaan berbasis pengetahuan dan analisis terhadap tanda-tanda yang ada. Prinsip ini sejalan dengan penerapan metode *Light Gradient Boosting Machine* (LightGBM) dalam penelitian ini, yang berfungsi untuk memprediksi jumlah pengunjung secara ilmiah dan terukur guna mendukung pengelolaan museum yang lebih efektif di masa mendatang.

Metode yang digunakan dalam penelitian ini adalah Light Gradient Boosting Machine (LightGBM), yaitu algoritma *machine learning* berbasis *Gradient Boosting Decision Tree (GBDT)*. LightGBM menggunakan strategi *leaf-wise growth* dengan pemilihan cabang berdasarkan pengurangan *loss* terbesar sehingga mampu menghasilkan model prediksi yang lebih optimal dibandingkan pendekatan *level-wise* pada algoritma boosting konvensional (Ayu et al., 2025). Keunggulan LightGBM terletak pada kecepatan pelatihan, efisiensi memori, serta

kemampuannya dalam menangani dataset besar dan fitur kategorikal tanpa perlu *encoding* tambahan (Fahmi et al., 2025). Dengan kelebihan tersebut, LightGBM banyak diaplikasikan dalam berbagai bidang prediksi, termasuk penelitian ini yang berfokus pada jumlah pengunjung museum.

Beberapa penelitian terdahulu telah membahas peramalan jumlah pengunjung museum dengan pendekatan metode yang berbeda. Penelitian oleh Najib & Nafi'iyah (2020) menggunakan algoritma Support Vector Machine (SVM) untuk memprediksi jumlah pengunjung museum di Jakarta. Hasil penelitian menunjukkan bahwa SVM mampu memberikan performa yang cukup baik dalam memodelkan data pengunjung, meskipun akurasinya masih dipengaruhi oleh variasi musiman dan tren tertentu. Sementara itu, penelitian oleh Yuliansyah (2025) mengusulkan kombinasi *Moving Average* dan *Prophet* untuk meramalkan jumlah pengunjung Museum Muhammadiyah. Hasil penelitian menunjukkan bahwa penggunaan *Moving Average* sebagai teknik *smoothing* berhasil meningkatkan akurasi model Prophet dengan nilai MAPE terbaik sebesar 5,58% pada data pengunjung personal harian, sehingga lebih stabil dalam menghadapi fluktuasi data. Kedua penelitian tersebut menegaskan pentingnya pemilihan metode yang tepat dalam menangani karakteristik data pengunjung museum yang cenderung fluktuatif.

Berdasarkan uraian tersebut, penelitian ini berkontribusi dalam mengembangkan sistem prediksi jumlah pengunjung harian Museum Mpu Purwa dengan menggunakan algoritma *Light Gradient Boosting Machine (LightGBM)*. Berbeda dengan penelitian terdahulu yang menggunakan SVM maupun kombinasi

Moving Average dan *Prophet*, penelitian ini menghadirkan pendekatan berbasis *machine learning* yang lebih efisien dalam pengolahan data berskala besar serta mampu mengatasi pola fluktuatif secara lebih adaptif (Liang et al., 2023). Kontribusi utama penelitian ini adalah memberikan model prediksi yang akurat dan dapat dijadikan dasar pengambilan keputusan strategis bagi pengelola museum, baik dalam aspek pengelolaan fasilitas, perencanaan kegiatan, maupun distribusi sumber daya manusia. Selain itu, penelitian ini juga diharapkan dapat menjadi referensi akademis dalam penerapan algoritma LightGBM pada bidang pariwisata dan pengelolaan destinasi budaya.

1.2 Rumusan Masalah

Bagaimana penerapan algoritma *Light Gradient Boosting Machine* (LightGBM) dalam menganalisis dan memprediksi jumlah pengunjung harian Museum MPU Purwa?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini ditetapkan agar ruang lingkup pembahasan lebih terarah dan fokus sesuai dengan tujuan penelitian. Adapun batasan masalah yang digunakan adalah sebagai berikut:

- a. Ruang lingkup penelitian ini berfokus pada prediksi jumlah pengunjung harian di Museum MPU Purwa.
- b. Algoritma yang digunakan terbatas pada *Light Gradient Boosting Machine* (LightGBM), tanpa membandingkan dengan metode lain.

- c. Dataset yang digunakan berasal dari data kunjungan harian Museum MPU Purwa tahun 2022–2025 yang diperoleh dari Dinas Pendidikan dan Kebudayaan Kota Malang.
- d. Faktor eksternal seperti kondisi cuaca, tren pariwisata global, maupun kegiatan eksternal di luar museum tidak dimasukkan dalam variabel penelitian.
- e. Evaluasi model dilakukan menggunakan metrik akurasi prediksi tertentu, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE).

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk memprediksi jumlah pengunjung harian Museum MPU Purwa menggunakan algoritma *Light Gradient Boosting Machine* (LightGBM) dengan tingkat akurasi yang memadai.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik dari sisi akademis maupun praktis. Adapun manfaat yang dimaksud adalah sebagai berikut:

- a. Memberikan kontribusi ilmiah dalam pengembangan kajian penerapan algoritma *Light Gradient Boosting Machine* (LightGBM) untuk prediksi jumlah pengunjung museum serta menjadi referensi bagi penelitian selanjutnya.
- b. Memberikan informasi prediksi jumlah pengunjung harian yang dapat dimanfaatkan oleh Museum MPU Purwa dan Dinas Pendidikan dan

Kebudayaan Kota Malang untuk mendukung pengelolaan fasilitas, penjadwalan kegiatan, dan strategi pengembangan museum.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terdahulu

Beberapa penelitian terdahulu yang berkaitan dengan forecasting dan prediksi jumlah pengunjung suatu tempat telah dilakukan oleh berbagai peneliti dengan beragam metode. Penelitian-penelitian tersebut menjadi dasar sekaligus referensi bagi penelitian ini karena memberikan gambaran mengenai pendekatan yang sudah diterapkan sebelumnya, baik menggunakan metode statistik maupun algoritma machine learning. Dengan menelaah penelitian terdahulu, dapat diketahui kelebihan, keterbatasan, serta peluang pengembangan yang masih terbuka. Oleh karena itu, hasil kajian penelitian sebelumnya sangat penting untuk memperkuat landasan teoretis sekaligus arah penelitian ini, di antaranya adalah:

Tabel 2.1 Penelitian Terdahulu

No	Nama Peneliti	Judul Penelitian	Metode	Hasil Penelitian
1	(Manullang et al., 2020)	Prediksi Kunjungan Taman Nasional Gunung Merbabu dengan Time Series Forecasting dan LSTM	<i>Time Series Forecasting</i> + LSTM	Model terbaik dengan tujuh lag, RMSE 3,7, efektif tingkatan akurasi prediksi.
2	(Cherry & Somya, 2023)	Prediksi Penjualan Tiket Wisata Taman Bermain menggunakan Metode ARIMA	ARIMA	ARIMA (1,0,0) signifikan, RMSE 21.296,39, mampu memprediksi tiket yang fluktuatif
3	(Iskandar, 2023)	Analisis Peramalan Jumlah Wisata Geopark Ciletuh Menggunakan <i>Exponential Smoothing</i>	<i>Exponential Smoothing</i>	MAPE 12,6%, mampu menangkap pola musiman dengan baik.
4	(Ramadhani et al., 2022)	Peramalan Jumlah Kunjungan Wisatawan Mancanegara ke Indonesia Menggunakan <i>Autoregressive Integrated Moving Average</i> (ARIMA)	ARIMA	ARIMA (2,1,1) akurat untuk data fluktuatif, terbatas pada pola non-linear kompleks

No	Nama Peneliti	Judul Penelitian	Metode	Hasil Penelitian
5	(Agustin et al., 2025)	Prediksi Jumlah Pengunjung Parawisata di Kabupaten Garut Menggunakan Algoritma Regresi Linear	Regresi Linear	Efektif untuk tren jangka panjang, tapi kurang optimal untuk pola musiman.
6	(Nulhakim et al., 2024)	Prediksi Jumlah Wisatawan Mancanegara ke Sumatera Utara Berdasarkan Pintu Masuk Utama Menggunakan Algoritma <i>Bacpropagation Neural Network</i>	Regresi Berganda	Akurat memprediksi berdasarkan pintu masuk bandara/pelabuhan.
7	(Budiawan et al., 2024)	Analisis Data Kunjungan Wisatawan Mancanegara ke Indonesia pada Era Pasca Pandemi melalui Metode Visualisasi dan Peramalan	Visualisasi + Peramalan	Menunjukkan tren penurunan pandemi lalu pemulihan di 2022.
8	(Wibowo & Tuti Purwaningsih, 2023)	Analisis Peramalan Jumlah Pengunjung Daya Tarik Wisata Kabupaten Bantul dengan Metode <i>Autoregressive Integrated Moving Average</i> (SARIMA)	SARIMA	SARIMA (1,1,1)(0,1,1) ₁₂ , MAPE 8,9%, akurat untuk pola musiman jelas.
Usulan Penelitian				
9	Prediksi Jumlah Pengunjung Harian Museum Mpu Purwa Menggunakan Metode LightGBM		LightGBM	Menghasilkan model prediksi jumlah pengunjung harian Museum Mpu Purwa dengan pendekatan LightGBM.

Pada penelitian (Manullang et al., 2020) prediksi kunjungan wisatawan di Taman Nasional Gunung Merbabu menggunakan *time series forecasting* dan *Long Short-Term Memory* (LSTM). Data yang digunakan adalah jumlah kunjungan bulanan 2013–2018. Dari 18 skenario yang diuji, model terbaik diperoleh dengan menggunakan tujuh *lag* sebagai fitur dan satu *lag* sebagai label, menghasilkan

RMSE terkecil sebesar 3,7. Hasil penelitian menunjukkan integrasi TSF dan LSTM efektif dalam menangkap pola deret waktu dan meningkatkan akurasi prediksi jumlah wisatawan.

Penelitian lain oleh (Cherry & Somya, 2023) memprediksi penjualan tiket wisata di Saloka Theme Park dengan metode *Autoregressive Integrated Moving Average* (ARIMA). Penelitian menggunakan data penjualan Januari 2020–Januari 2023. Model terbaik ARIMA (1,0,0) signifikan secara statistik ($p\text{-value} = 0.001 < 0.05$) dan menghasilkan nilai RMSE 21.296,39. Prediksi menunjukkan peningkatan penjualan pada Februari–Maret 2023 masing-masing sebesar 28.983 dan 84.931 tiket. Penelitian ini menyimpulkan bahwa ARIMA mampu memodelkan data penjualan tiket yang bersifat fluktuatif secara cukup akurat.

Penelitian oleh (Iskandar, 2023) membahas peramalan jumlah wisatawan Geopark Ciletuh menggunakan metode *Exponential Smoothing*. Data yang dianalisis adalah jumlah kunjungan wisatawan bulanan pada periode 2018–2021. Hasil penelitian menunjukkan bahwa model *Exponential Smoothing* mampu menangkap pola musiman dalam data kunjungan dengan cukup baik, ditunjukkan oleh nilai *Mean Absolute Percentage Error* (MAPE) sebesar 12,6%. Kesimpulannya, metode ini dapat digunakan sebagai dasar dalam membantu pengelolaan destinasi wisata berbasis data kunjungan aktual.

Penelitian lain oleh (Ramadhani et al., 2022) memprediksi jumlah kunjungan wisatawan mancanegara ke Indonesia menggunakan metode *Autoregressive Integrated Moving Average* (ARIMA). Data yang digunakan adalah kunjungan wisatawan bulanan dari Badan Pusat Statistik (BPS) tahun 2015–2020.

Model terbaik yang diperoleh adalah ARIMA (2,1,1) dengan tingkat akurasi prediksi yang cukup tinggi. Penelitian ini menyimpulkan bahwa ARIMA mampu memodelkan data kunjungan yang fluktuatif, meskipun keterbatasan muncul saat data memiliki pola non-linear yang kompleks.

Penelitian oleh (Agustin et al., 2025) melakukan prediksi jumlah pengunjung pariwisata di Kabupaten Garut menggunakan algoritma regresi linear. Data yang digunakan adalah kunjungan wisatawan tahunan periode 2010–2019. Hasil penelitian menunjukkan bahwa regresi linear dapat memberikan gambaran tren peningkatan jumlah wisatawan, meskipun akurasinya terbatas karena metode ini kurang mampu menangkap pola musiman. Kesimpulannya, regresi linear efektif untuk melihat tren jangka panjang, tetapi kurang optimal untuk data dengan fluktuasi tinggi.

Sementara itu, (Nulhakim et al., 2024) meneliti prediksi jumlah wisatawan mancanegara ke Sumatera Utara berdasarkan pintu masuk kedatangan. Metode yang digunakan adalah analisis regresi berganda dengan data dari BPS periode 2015–2019. Hasil penelitian menunjukkan bahwa jumlah kunjungan wisatawan dipengaruhi secara signifikan oleh faktor pintu masuk bandara dan pelabuhan, dengan tingkat akurasi prediksi yang cukup baik. Kesimpulannya, regresi berganda dapat digunakan untuk memprediksi kunjungan wisatawan, terutama jika terdapat variabel independen tambahan seperti pintu masuk kedatangan.

Penelitian oleh (Budiawan et al., 2024) membahas analisis data kunjungan wisatawan mancanegara ke Indonesia pada era pasca pandemi dengan metode visualisasi dan peramalan. Data yang digunakan berasal dari BPS periode 2019–

2022. Hasil penelitian menunjukkan adanya penurunan drastis jumlah kunjungan pada tahun 2020 akibat pandemi COVID-19, lalu peningkatan kembali pada 2022. Dengan metode visualisasi, pola penurunan dan kenaikan lebih mudah diamati, sedangkan hasil peramalan menunjukkan tren positif pemulihan kunjungan wisatawan. Kesimpulannya, metode visualisasi dipadukan dengan peramalan mampu memberikan gambaran komprehensif mengenai dinamika kunjungan pasca pandemi.

Selanjutnya, (Wibowo & Tuti Purwaningsih, 2023) melakukan analisis peramalan jumlah pengunjung daya tarik wisata di Kabupaten Bantul menggunakan metode *Seasonal Autoregressive Integrated Moving Average* (SARIMA). Data yang digunakan berupa jumlah pengunjung bulanan periode 2017–2021. Model terbaik yang diperoleh adalah SARIMA (1,1,1)(0,1,1)₁₂ dengan nilai MAPE sebesar 8,9%. Hasil penelitian menunjukkan model ini mampu menangkap pola musiman kunjungan dengan akurasi yang tinggi. Kesimpulannya, metode SARIMA efektif untuk memprediksi jumlah kunjungan wisata yang memiliki tren musiman jelas.

2.2 Time Series dan Forecasting

Time series atau deret waktu merupakan data yang dicatat secara berurutan berdasarkan interval tertentu dan memiliki ketergantungan antarperiode (Ardesfira et al., 2022). Data ini umumnya dianalisis untuk melihat pola yang terbentuk dari waktu ke waktu. Komponen penting dalam analisis deret waktu meliputi *trend*, *seasonality*, *cyclical*, dan *residual* yang membantu mengidentifikasi pola

tersembunyi dalam data. Pemahaman terhadap komponen tersebut sangat penting agar hasil analisis dan peramalan dapat lebih akurat.

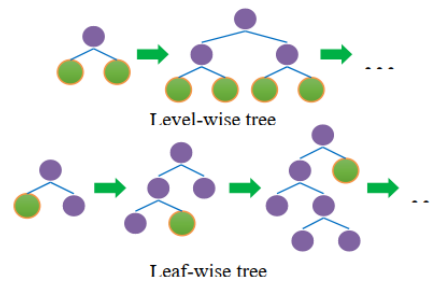
Forecasting atau peramalan adalah proses memperkirakan nilai di masa mendatang dengan memanfaatkan pola historis pada data deret waktu (Arwansyah et al., 2022). Peramalan dapat dilakukan dengan pendekatan kualitatif maupun kuantitatif tergantung pada kebutuhan penelitian. Metode kuantitatif biasanya menggunakan model matematis atau algoritma, seperti ARIMA, SARIMA, LightGBM, atau GRU. Hasil dari peramalan dapat digunakan untuk mendukung pengambilan keputusan, misalnya dalam mengantisipasi lonjakan atau penurunan pengunjung.

2.3 Metode *LightGBM*

Light Gradient Boosting Machine (LightGBM) adalah algoritma *machine learning* berbasis *Gradient Boosting Decision Tree* (GBDT) yang dikembangkan oleh *Microsoft*. Algoritma ini dirancang untuk menghasilkan prediksi dengan akurasi tinggi namun tetap efisien dalam hal waktu komputasi dan penggunaan memori. LightGBM menggunakan konsep *ensemble learning*, yaitu menggabungkan banyak pohon keputusan (*decision tree*) secara bertahap untuk memperbaiki kesalahan prediksi pada iterasi sebelumnya. Selain itu, LightGBM dikenal sebagai metode *gradient boosting* yang cepat, terdistribusi, serta mampu mengatasi data berskala besar dengan kebutuhan memori rendah (Kurniadi & Larasati, 2022)

Prinsip utama yang membedakan LightGBM dengan GBDT konvensional adalah strategi pembangunan pohon yang disebut *leaf-wise growth*. Pada metode

konvensional, pohon dibangun dengan pendekatan *level-wise growth* di mana semua node pada level yang sama di *split* secara bersamaan. Sebaliknya, LightGBM memilih node (*leaf*) dengan nilai penurunan error terbesar untuk diperluas terlebih dahulu. Strategi ini membuat pohon dapat tumbuh lebih dalam pada bagian yang paling informatif sehingga lebih cepat menemukan pola kompleks dalam data. Namun, *leaf-wise growth* juga dapat meningkatkan risiko *overfitting* apabila kedalaman pohon tidak dibatasi dengan baik. Perbedaan strategi pertumbuhan pohon ini dapat dilihat pada Gambar 2.1 yang menunjukkan perbandingan antara konstruksi *level-wise* dan *leaf-wise* (Septiana Rizky et al., 2022)



Gambar 2.1 Perbedaan *Level-wise tree* dan *Leaf-wise tree*

Secara umum, alur kerja LightGBM dimulai dengan meminimalkan fungsi kehilangan (*loss function*), yang secara umum dapat ditulis sebagai:

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2.1)$$

Keterangan :

- L : total *loss function* (fungsi kerugian) yang ingin diminimalkan.
- n : jumlah data sampel
- y_i : nilai aktual (data asli) pada sampel ke- i
- $\hat{y}_i$: nilai prediksi dari model pada sampel ke- i
- $l(y_i, \hat{y}_i)$: fungsi kerugian individual, misalnya selisih kuadrat $(y_i - \hat{y}_i)^2$
- K : jumlah pohon (trees) yang digunakan dalam boosting.
- $\Omega(f_k)$: istilah regulasi (*regularization*) untuk mengontrol kompleksitas pohon ke- k

Untuk kasus *Mean Squared Error (MSE)*, gradien dan hessian diturunkan menjadi:

$$g_i = \hat{y}_i - y_i, h_i = 1 \quad (2.2)$$

Keterangan :

- g_i : total *loss function* (fungsi kerugian) yang ingin diminimalkan.
- h_i : hessian pada sampel ke- i
- $\hat{y}_i$: nilai prediksi dari model
- y_i : nilai aktual (data asli)

Nilai keluaran optimal dari sebuah daun dihitung dengan:

$$w^* = - \frac{\sum_{i \in I} g_i}{\sum_{i \in I} h_i + \lambda} \quad (2.3)$$

Keterangan :

- w^* : nilai keluaran optimal untuk sebuah daun (leaf) dalam pohon keputusan.
- I : himpunan data sampel yang masuk ke leaf tersebut.
- g_i : gradien dari fungsi loss terhadap prediksi $\hat{y}_i$. Biasanya $g_i = \hat{y}_i - y_i$.
- h_i : hessian, yaitu turunan kedua fungsi loss (ukuran kelengkungan)
- λ : parameter regulasi untuk mengendalikan besar kecilnya bobot agar tidak terlalu ekstrem.

Sementara itu, *gain* atau keuntungan dari split node dihitung sebagai:

$$Gain = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right) - \gamma \quad (2.4)$$

Keterangan :

- Gain* : ukuran seberapa besar penurunan error jika sebuah node dibagi (split).
- G_L, G_R : jumlah gradien pada cabang kiri dan kanan setelah split.
- H_L, H_R : jumlah hessian pada cabang kiri dan kanan.
- λ : parameter regulasi bobot.
- γ : parameter regularisasi tambahan, memberi penalti setiap kali membuat split baru.

Dengan prinsip ini, LightGBM bekerja membangun pohon secara iteratif.

Proses dimulai dari data masukan yang terlebih dahulu diubah ke dalam bentuk

histogram agar pemrosesan lebih efisien. Selanjutnya dilakukan optimasi melalui *Exclusive Feature Bundling (EFB)* yang menggabungkan fitur jarang muncul serta *Gradient-based One-Side Sampling (GOSS)* yang memprioritaskan data dengan gradien besar. Setelah tahap tersebut, pohon dibangun dengan strategi *leaf-wise growth* sehingga percabangan diarahkan pada bagian yang paling memberikan penurunan error. Proses ini dilakukan berulang hingga diperoleh model akhir yang siap digunakan untuk melakukan prediksi.

Dalam penelitian ini, fungsi kerugian yang digunakan adalah *Mean Squared Error* (MSE) karena sesuai untuk kasus regresi. MSE dihitung sebagai rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi, dengan rumus:

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

Keterangan :

MSE : rata-rata kuadrat selisih antara nilai aktual dan prediksi.

n : jumlah data

y_i : nilai aktual (data asli) pada sampel ke- i

$\hat{y}_i$: nilai prediksi dari model pada sampel ke- i

Dengan demikian, penggunaan MSE sebagai fungsi kerugian dalam penelitian ini bertujuan untuk meminimalkan selisih antara nilai aktual dan nilai prediksi pada setiap iterasi pembangunan pohon. Semakin kecil nilai MSE yang dihasilkan, maka semakin baik pula model dalam melakukan prediksi. Oleh karena itu, MSE tidak hanya digunakan sebagai acuan dalam proses pelatihan model, tetapi juga menjadi tolok ukur awal untuk membandingkan perbaikan error pada setiap pohon sebelum dilanjutkan ke tahap evaluasi akhir dengan metrik lain seperti MAE, RMSE, dan MAPE.

2.4 Feature Engineering

Feature engineering merupakan proses pengolahan data untuk menghasilkan variabel baru atau memodifikasi variabel yang sudah ada agar lebih informatif bagi model prediksi. Tahap ini penting karena kualitas fitur akan sangat memengaruhi akurasi model dalam mengenali pola dari data. Penelitian terdahulu menunjukkan bahwa penerapan feature engineering, seperti oversampling, standardisasi, dan seleksi fitur, dapat meningkatkan kinerja model secara signifikan. Misalnya, pada penelitian klasifikasi risiko kredit, akurasi model *Random Forest* meningkat dari 92,56% menjadi 97,94% setelah dilakukan *feature engineering* berupa *oversampling* dan standardisasi, sementara tuning parameter tidak memberikan peningkatan yang signifikan (Nur Fauzi et al., 2025).

Salah satu teknik yang umum digunakan adalah pembuatan *lag features*, yaitu menambahkan variabel baru yang merepresentasikan nilai pada periode sebelumnya. Misalnya, jumlah pengunjung pada hari ke- $t - 1$ dapat didefinisikan sebagai:

$$X_{t-1} = \text{Jumlah pengunjung pada hari sebelumnya} \quad (2.6)$$

Keterangan :

X_{t-1} : nilai jumlah pengunjung pada hari ke- $t - 1$

t : indeks waktu saat ini

Selain itu, digunakan juga teknik *rolling window* atau *moving average* untuk memperhalus fluktuasi data dan menangkap tren jangka pendek. Moving average dengan jendela sepanjang n periode dapat dituliskan sebagai:

$$MA_t = \frac{1}{n} \sum_{i=0}^{n-1} X_{t-i} \quad (2.7)$$

Keterangan :

MA_t : nilai rata-rata bergerak pada waktu ke- t

n : panjang jendela (jumlah periode yang dirata-ratakan).

X_{t-i} : jumlah pengunjung pada periode ke- $t - i$

Dengan penerapan feature engineering seperti ekstraksi waktu (hari, bulan, tahun, akhir pekan, atau hari libur), lag features, dan rolling window, model LightGBM dapat lebih mudah memahami pola yang kompleks pada data jumlah pengunjung. Hal ini pada akhirnya meningkatkan performa prediksi sekaligus menjaga efisiensi pemodelan.

2.5 Evaluasi Model

Evaluasi model merupakan tahap penting dalam penelitian prediksi untuk menilai sejauh mana kinerja metode yang digunakan dalam menghasilkan hasil yang akurat. Tanpa adanya evaluasi, sulit untuk mengetahui apakah model yang dibangun sudah mampu merepresentasikan pola data dengan baik atau masih memerlukan perbaikan (Saputra & Sugihartono, 2025).

Dalam penelitian deret waktu, terdapat beberapa ukuran yang umum digunakan untuk menilai performa model. Salah satunya adalah *Mean Absolute Error* (MAE), yaitu rata-rata nilai absolut selisih antara data aktual dengan hasil prediksi:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.8)$$

Keterangan :

MAE : rata-rata selisih absolut antara nilai aktual dan nilai prediksi.

n : jumlah data

y_i : nilai aktual (data asli) pada sampel ke- i

$\hat{y}_i$: nilai prediksi dari model pada sampel ke- i

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.9)$$

Keterangan :

$RMSE$: akar kuadrat dari MSE.

n : jumlah data

y_i : nilai aktual (data asli) pada sampel ke- i

$\hat{y}_i$: nilai prediksi dari model pada sampel ke- i

Selain itu, *Mean Absolute Percentage Error* (MAPE) juga sering dipakai karena memberikan ukuran kesalahan dalam bentuk persentase, sehingga lebih mudah diinterpretasikan:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2.10)$$

Keterangan :

$MAPE$: rata-rata persentase kesalahan prediksi terhadap nilai aktual.

n : jumlah data

y_i : nilai aktual (data asli) pada sampel ke- i

$\hat{y}_i$: nilai prediksi dari model pada sampel ke- i

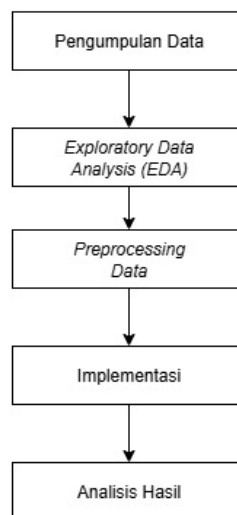
Dengan menggunakan metrik-metrik tersebut, peneliti dapat membandingkan performa antar model serta menentukan metode yang paling sesuai untuk digunakan dalam peramalan jumlah pengunjung museum.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian menjelaskan alur yang ditempuh dalam penelitian ini untuk mencapai tujuan yang telah ditetapkan. Proses penelitian disusun secara sistematis dalam bentuk blok diagram sehingga setiap tahap dapat dipahami dengan jelas.



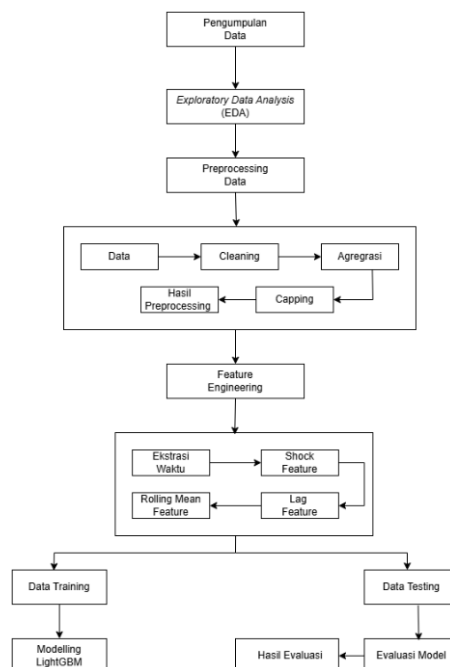
Gambar 3.1 Desain Penelitian

Secara garis besar, penelitian ini dimulai dari pengumpulan data kunjungan Museum Mpu Purwa. Selanjutnya dilakukan *exploratory data analysis* (EDA) untuk memahami karakteristik dataset, termasuk tren kunjungan, pola musiman, serta potensi adanya outlier. Setelah itu, data diproses melalui tahap preprocessing yang meliputi pembersihan, agregasi, serta transformasi agar sesuai dengan kebutuhan pemodelan.

Data yang telah diproses kemudian digunakan pada tahap implementasi model, di mana algoritma *Light Gradient Boosting Machine* (LightGBM) diterapkan untuk melakukan prediksi jumlah pengunjung harian. Tahap akhir adalah analisis hasil, yaitu mengevaluasi performa model menggunakan metrik error seperti MAE, RMSE, dan MAPE untuk mengukur tingkat akurasi prediksi yang dihasilkan.

3.2 Desain Sistem

Desain sistem menjelaskan alur yang ditempuh dalam penelitian ini untuk mencapai tujuan yang telah ditetapkan. Proses perancangan sistem disusun secara sistematis dalam bentuk blok diagram sehingga setiap tahap dapat dipahami dengan jelas, mulai dari proses awal pengolahan data hingga model selesai dievaluasi.



Gambar 3.2 Desain Sistem

Secara garis besar, sistem yang dibangun terdiri dari beberapa tahap utama. Tahap pertama adalah *preprocessing data*, yang mencakup proses *cleaning*, agregasi, dan *capping* untuk memastikan data siap digunakan. Tahap selanjutnya adalah *feature engineering*, di mana dilakukan ekstraksi fitur tambahan seperti informasi waktu (hari, bulan), pembuatan fitur shock, pembuatan lag feature, serta rolling/moving statistic untuk memperkaya variabel input. Data yang telah diproses kemudian dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk membangun model prediksi menggunakan algoritma *LightGBM*, sementara data *testing* digunakan untuk mengevaluasi performa model. Hasil evaluasi kemudian dianalisis untuk mengetahui sejauh mana model mampu memprediksi jumlah pengunjung dengan akurat.

3.2.1 Exploratory Data Analysis (EDA)

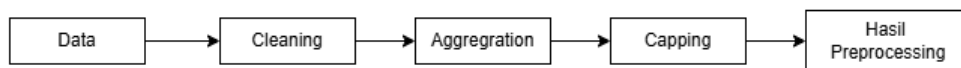
Exploratory Data Analysis (EDA) merupakan tahap analisis awal yang bertujuan untuk memahami karakteristik data yang digunakan dalam penelitian. Pada tahap ini dilakukan eksplorasi untuk mengidentifikasi pola umum, distribusi data, nilai hilang (missing values), serta adanya data yang tidak wajar (outlier). Proses ini penting agar peneliti dapat memperoleh gambaran awal mengenai kualitas data dan potensi permasalahan yang dapat memengaruhi hasil pemodelan.

Melalui EDA, peneliti dapat mengetahui jumlah data yang tersedia, rentang periode data, serta variasi jumlah pengunjung per hari. Selain itu, EDA juga membantu mengidentifikasi apakah terdapat tren musiman atau pola tertentu yang muncul dari data pengunjung museum. Hasil eksplorasi ini tidak hanya menjadi dasar dalam pengambilan keputusan untuk tahap praproses data selanjutnya, tetapi

juga menjadi bahan pertimbangan dalam merancang fitur tambahan pada tahap *feature engineering*.

3.2.2 Preprocessing Data

Tahap *preprocessing* data dilakukan untuk mempersiapkan data sebelum digunakan pada proses pemodelan. Proses ini meliputi beberapa langkah penting, yaitu *data cleaning* untuk mengatasi data yang tidak valid atau hilang, *data transformation* untuk mengubah format atau skala data agar sesuai kebutuhan analisis, serta *data aggregation* untuk merangkum data berdasarkan periode waktu tertentu. Dengan adanya tahap ini, data menjadi lebih bersih, konsisten, dan siap diproses lebih lanjut. Seperti yang ditunjukkan pada Gambar 3.3, berikut merupakan tahapan preprocessing data.



Gambar 3.3 Alur Kerja *Preprocessing Data*

3.2.2.1 Cleaning Data

Data cleaning merupakan tahap penting dalam *preprocessing* untuk memastikan data yang digunakan berkualitas baik, konsisten, dan relevan dengan tujuan penelitian. Dari total 18 atribut yang tersedia, tidak semuanya digunakan dalam proses prediksi. Tahap ini tidak hanya mencakup pembersihan nilai yang tidak wajar (*outlier*), tetapi juga pemilihan fitur (*feature selection*). Berdasarkan tujuan penelitian, hanya dua atribut yang dipertahankan, yaitu *timestamp* sebagai penanda waktu kunjungan dan jumlah pengunjung sebagai variabel target yang diprediksi.

Tabel 3.1 Sebelum *Cleaning Data*

Timestamp	Nama	Asal	Alamat	Jumlah Pengunjung
2/5/2023 10:34:02	Azzaria Selena	SMKN 7 MALANG	...	6
2/5/2023 10:51:48	Elvina Rinaldy	SMAK KOLESE SANTO YUSUP	...	9
2/5/2023 10:57:37	Kiera	SMAK Kolese Santo Yusup Malang	...	5
2/5/2023 11:55:09	Rizka Aulia Destani	SMPN 3 Malang	...	5
2/7/2023 9:03:25	Putri Nur	SMPN 5 SINGOSARI SATU ATAP	...	100
2/7/2023 9:05:07	Moh. Shalhan (pendamping)	SMPN 5 SINGOSARI SATU ATAP	...	100
2/7/2023 12:37:23	Mohammad Alfi Hidayah	SDN Purwantoro 2 Malang	...	13
2/8/2023 14:23:19	Adi Netto	UKDW	...	2
2/9/2023 11:30:42	Nandya	SD NEGERI MODEL KOTA MALANG	...	127
2/10/2023 9:52:55	Dewi	UB	...	1

Setelah dilakukan *cleaning*, atribut yang tidak relevan dengan proses prediksi seperti nama pengunjung, asal instansi, alamat, dan kolom-kolom lainnya yang tidak dibutuhkan dihapus. Pembersihan juga dilakukan terhadap nilai jumlah pengunjung yang dianggap outlier, misalnya entri dengan 100 atau 127 pengunjung, karena secara distribusi data jumlah pengunjung harian umumnya berada di bawah 100 orang. Dengan demikian, hanya atribut *timestamp* dan jumlah pengunjung yang dipertahankan untuk tahap analisis berikutnya.

Tabel 3.2 Setelah *Cleaning Data*

Timestamp	Jumlah Pengunjung	Keterangan
2/5/2023 10:34:02	6	Valid
2/5/2023 10:51:48	9	Valid
2/5/2023 10:57:37	5	Valid
2/5/2023 11:55:09	5	Valid
2/7/2023 9:03:25	100	Outlier
2/7/2023 9:05:07	100	Outlier
2/7/2023 12:37:23	13	Valid
2/8/2023 14:23:19	2	Valid
2/9/2023 11:30:42	127	Outlier
2/10/2023 9:52:55	1	Valid

3.2.2.2 Data Aggregation

Data aggregation merupakan tahap pengolahan data dengan cara menggabungkan atau merangkum data mentah ke dalam satuan tertentu agar lebih mudah dianalisis. Pada penelitian ini, data kunjungan museum awalnya dicatat per entri kunjungan, di mana setiap baris merepresentasikan satu instansi atau individu pada waktu tertentu. Agar sesuai dengan tujuan penelitian, data tersebut kemudian diagregasikan menjadi jumlah pengunjung per hari. Dengan cara ini, pola kunjungan harian dapat dianalisis secara lebih jelas dan dijadikan input bagi model prediksi.

Tabel 3.3 Sebelum *Agregation Data*

Timestamp	Jumlah Pengunjung
2/5/2023 10:34:02	6
2/5/2023 10:51:48	9
2/5/2023 10:57:37	5
2/5/2023 11:55:09	5
2/7/2023 12:37:23	13
2/8/2023 14:23:19	2
2/9/2023 11:30:42	0
2/10/2023 9:52:55	1

Proses agregasi dilakukan dengan menjumlahkan seluruh nilai pengunjung pada tanggal yang sama menggunakan rumus:

$$\text{Total Pengunjung Harian} = \sum_{i=1}^n x_i \quad (3.1)$$

Keterangan :

n : jumlah entri kunjungan pada suatu hari tertentu

i : indeks kunjungan ke- i pada hari tersebut

x_i : jumlah pengunjung pada kunjungan ke- i

Dengan sampel data sebelumnya, salah satu perhitungan yang diperoleh contohnya pada tanggal 5 Februari 2023 terdapat 4 entri kunjungan dengan jumlah

pengunjung masing-masing 6, 9, 5, dan 5 orang. Maka total pengunjung harian adalah:

$$Total\ Pengunjung\ Harian = \sum_{i=1}^n x_i = 6 + 9 + 5 + 5 = 25$$

Sehingga, dapat disimpulkan bahwa pada tanggal 5 Februari 2023 terdapat 25 pengunjung. Perhitungan data yang telah diagregasi secara lengkap dapat dilihat pada Tabel 3.6.

Tabel 3.4 Setelah *Agregation Data*

Timestamp	Jumlah Pengunjung
2/5/2023	25
2/7/2023	13
2/8/2023	2
2/9/2023	0
2/10/2023	1

3.2.2.3 *Capping*

Capping merupakan teknik penanganan outlier dengan cara membatasi nilai ekstrem berdasarkan batas statistik tertentu tanpa menghapus data tersebut. Pada penelitian ini, proses capping dilakukan menggunakan metode *Interquartile Range* (IQR) untuk menentukan batas atas nilai yang dianggap sebagai outlier.

Metode IQR bekerja dengan menghitung kuartil pertama (Q_1) dan kuartil ketiga (Q_3) dari distribusi data jumlah pengunjung harian. Selisih antara Q_3 dan Q_1 disebut sebagai IQR, yang dirumuskan sebagai berikut:

$$IQR = Q_3 - Q_1 \quad (3.2)$$

Batas atas (upper bound) untuk mendeteksi outlier dihitung dengan rumus:

$$Batas\ Atas = Q_3 + 1.5 \times IQR \quad (3.3)$$

Sedangkan batas bawah (lower bound) dihitung dengan:

$$\text{Batas Bawah} = Q_3 - 1.5 \times IQR \quad (3.4)$$

Pada penelitian ini, fokus capping dilakukan pada nilai yang melebihi batas atas. Setiap nilai jumlah pengunjung harian yang lebih besar dari batas atas akan disesuaikan menjadi sebesar nilai batas atas tersebut. Secara matematis, proses capping dapat dituliskan sebagai:

$$x_{capped} = \begin{cases} \text{Batas Atas, jika } x > \text{Batas Atas} \\ x, & \text{jika } x \leq \text{Batas Atas} \end{cases}$$

Keterangan :

x : jumlah pengunjung harian sebelum dicapping

x_{capped} : jumlah pengunjung harian setelah dicapping

Dengan menggunakan metode IQR, proses *capping* menjadi lebih objektif karena batas ditentukan berdasarkan distribusi data aktual, bukan berdasarkan nilai asumsi tertentu. Teknik ini membantu mengurangi pengaruh nilai ekstrem terhadap proses pelatihan model LightGBM tanpa menghilangkan data dari dataset.

Tabel 3.5 Setelah Proses *Capping*

Tanggal	Jumlah Pengunjung	Batas Atas (IQR)	Jumlah Setelah Capping
2/9/2023	127	58	58
2/7/2023	100	58	58
2/5/2023	25	58	25

3.2.3 Feature Engineering

Feature engineering merupakan tahap pembuatan variabel baru yang relevan untuk meningkatkan performa model prediksi. Pada penelitian ini, data yang sudah melalui tahap preprocessing diperkaya dengan beberapa fitur tambahan. Penambahan fitur ini dilakukan untuk membantu algoritma *LightGBM* mengenali pola dan hubungan dalam data secara lebih optimal.

3.2.3.1 Ekstraksi Fitur Waktu

Ekstraksi waktu dilakukan dengan memanfaatkan atribut tanggal hasil agregasi. Dari atribut ini diturunkan beberapa variabel baru yang bertujuan untuk menangkap pola musiman maupun periodik kunjungan museum. Variabel yang dihasilkan antara lain *day*, *month*, *year*, *day_of_week*, serta *is_weekend*. Rumus penurunan variabel ditunjukkan sebagai berikut:

$$\begin{aligned} day &= DAY(date), month = MONTH(date), year = YEAR(date) \\ day_{of_week} &= WEEKDAY(date), is_{weekend} = \begin{cases} 1, & \text{jika } day_{of_week} \in \{5,6\} \\ 0, & \text{lainnya} \end{cases} \end{aligned} \quad (3.5)$$

Keterangan :

day : tanggal dalam bulan (1-31)

month : bulan dalam tahun (1-12)

year : tahun pengamatan

day_{of_week} : nilai 0-6 yang merepresentasikan hari Senin – Minggu

is_{weekend} : variabel biner, jika bernilai 1 merupakan weekend (hari Sabtu / Minggu) dan bernilai 0 jika merupakan weekday (hari Senin – Minggu)

Sebagai contoh, tanggal 5 Februari 2023 dikonversi menjadi hari Minggu, bulan Februari, tahun 2023, dengan *day_of_week* = 6, dan dikategorikan sebagai akhir pekan (*is_weekend* = 1). Sehingga, diperoleh hasil seperti pada tabel 3.8.

Tabel 3.6 Tabel data setelah ekstraksi fitur waktu

Tanggal	Total Pengunjung	Hari	Bulan	Tahun	<i>day_of_week</i>	<i>is_weekend</i>
2/5/2023	25	Minggu	Februari	2023	6	1
2/7/2023	13	Selasa	Februari	2023	1	0
2/8/2023	2	Rabu	Februari	2023	2	0
2/9/2023	0	Kamis	Februari	2023	3	0
2/10/2023	1	Jum'at	Februari	2023	4	0

3.2.3.2 Shock Feature

Shock feature merupakan variabel yang digunakan untuk menangkap perubahan atau lonjakan jumlah pengunjung antar hari. Fitur ini bertujuan untuk

membantu model mengenali adanya kenaikan atau penurunan yang signifikan dibandingkan hari sebelumnya. Dengan memasukkan informasi perubahan tersebut, model dapat memahami dinamika fluktuasi jumlah pengunjung secara lebih baik.

Pada penelitian ini, shock feature dihitung sebagai selisih antara jumlah pengunjung pada hari ke-t dengan jumlah pengunjung pada hari sebelumnya (t-1).

Secara matematis dapat dirumuskan sebagai berikut:

$$Shock_t = y_t - y_{t-1} \quad (3.6)$$

Keterangan :

$Shock_t$: nilai perubahan jumlah pengunjung pada hari ke-t

y_t : jumlah pengunjung pada hari ke-t

y_{t-1} : jumlah pengunjung pada hari sebelumnya

Jika nilai shock bernilai positif, maka terjadi peningkatan jumlah pengunjung dibandingkan hari sebelumnya. Sebaliknya, jika bernilai negatif, maka terjadi penurunan jumlah pengunjung.

Sebagai contoh, jumlah pengunjung pada tanggal 12 Februari 2023 adalah 18 orang, sedangkan pada tanggal 11 Februari 2023 adalah 51 orang. Maka nilai shock dihitung sebagai berikut:

$$Shock_{12} = 18 - 51 = -33$$

Nilai -33 menunjukkan bahwa terjadi penurunan jumlah pengunjung sebanyak 33 orang dibandingkan hari sebelumnya. Dengan adanya shock feature, model LightGBM tidak hanya melihat jumlah pengunjung secara absolut, tetapi juga mempertimbangkan perubahan antar waktu sehingga pola lonjakan dan penurunan dapat dipelajari dengan lebih efektif.

3.2.3.3 Lag Feature

Lag feature dibentuk untuk menangkap pengaruh jumlah kunjungan pada periode sebelumnya terhadap hari berjalan. Beberapa fitur yang digunakan adalah *lag-1* yang merupakan kunjungan 1 hari sebelumnya dan *lag-7* yang merupakan kunjungan 7 hari sebelumnya.

Penggunaan *lag feature* ini penting karena pola jumlah pengunjung di suatu hari biasanya dipengaruhi oleh hari-hari sebelumnya. Misalnya, jika beberapa hari terakhir ramai, kemungkinan besar hari berikutnya juga cenderung ramai, atau sebaliknya. Dengan memasukkan informasi dari periode sebelumnya, model *LightGBM* dapat belajar hubungan *time dependency* (ketergantungan waktu) sehingga prediksi menjadi lebih akurat dibandingkan hanya mengandalkan data pada hari itu saja.

$$lag_k(t) = y_{t-k} \quad (3.7)$$

Keterangan :

$lag_k(t)$: nilai jumlah pengunjung pada hari ke- $t - k$

Sebagai ilustrasi, jumlah pengunjung pada tanggal 12 Februari 2023 adalah 18 orang. Dengan menggunakan *lag-1*, maka nilai yang digunakan adalah jumlah pengunjung pada tanggal 11 Februari 2023 yaitu 51 orang.

$$lag_1(12) = y_{12-1} = y_{11} = 51$$

Sedangkan untuk *lag-7*, nilai yang digunakan adalah jumlah pengunjung pada tanggal 5 Februari 2023 yaitu 25 orang.

$$lag_7(12) = y_{12-7} = y_5 = 25$$

Sehingga, hasil data yang sudah mendapatkan fitur ekstrasi waktu seperti pada tabel 3.7:

Tabel 3.7 Tabel Data setelah *Lag Feature*

Tanggal	Total Pengunjung	<i>lag-1</i>	<i>lag-7</i>
2/5/2023	25	...	...
2/7/2023	13	...	...
2/8/2023	2	13	...
2/9/2023	0	2	...
2/10/2023	1	2	...
2/11/2023	51	1	...
2/12/2023	18	51	25

3.2.3.4 *Rolling Moving / Statistic*

Rolling moving / statistic digunakan untuk menangkap tren jangka pendek maupun jangka panjang dengan menghitung rata-rata dari beberapa periode sebelumnya. Pada penelitian ini digunakan *rolling mean* 7 hari untuk merepresentasikan pola mingguan (*weekday-weekend*) serta *rolling mean* 30 hari untuk menggambarkan kecenderungan bulanan yang lebih stabil. Dengan cara ini, model dapat lebih mudah memahami apakah jumlah pengunjung sedang mengalami kenaikan, penurunan, atau berada pada kondisi relatif stabil.

$$MA_t = \frac{1}{n} \sum_{i=0}^{n-1} X_{t-i} \quad (3.8)$$

Keterangan :

MA_t : nilai rata-rata bergerak pada waktu ke- t

n : panjang jendela (jumlah periode yang dirata-ratakan).

X_{t-i} : jumlah pengunjung pada periode ke- $t - i$

Sebagai contoh, jumlah pengunjung pada tanggal 5 Februari 2023, 6 Februari 2023, 7 Februari 2023, 8 Februari 2023, 9 Februari 2023, 10 Februari 2023, dan 11 Februari 2023 berturut-turut adalah 25, 0, 13, 2, 0, 51, dan 18. Dengan

menggunakan *rolling mean* 7, maka rata-rata jumlah pengunjung pada tanggal 11 Februari 2023 dihitung sebagai:

$$MA_{11} = \frac{1}{7} (X_{11} + X_{10} + X_9 + X_8 + X_7 + X_6 + X_5)$$

$$MA_{11} = \frac{1}{7} (18 + 51 + 0 + 2 + 13 + 0 + 25) \approx 15.57$$

Sehingga, *rolling mean* 7 untuk rata-rata pengunjung pada tanggal 11 Februari menghasilkan nilai sebesar 15,57, yang menggambarkan kecenderungan jumlah pengunjung dalam jangka pendek (mingguan). Dengan prinsip perhitungan yang sama, *rolling mean* 30 menggunakan jendela 30 hari untuk melihat tren jangka panjang (bulanan).

3.2.4 Modelling *LightGBM*

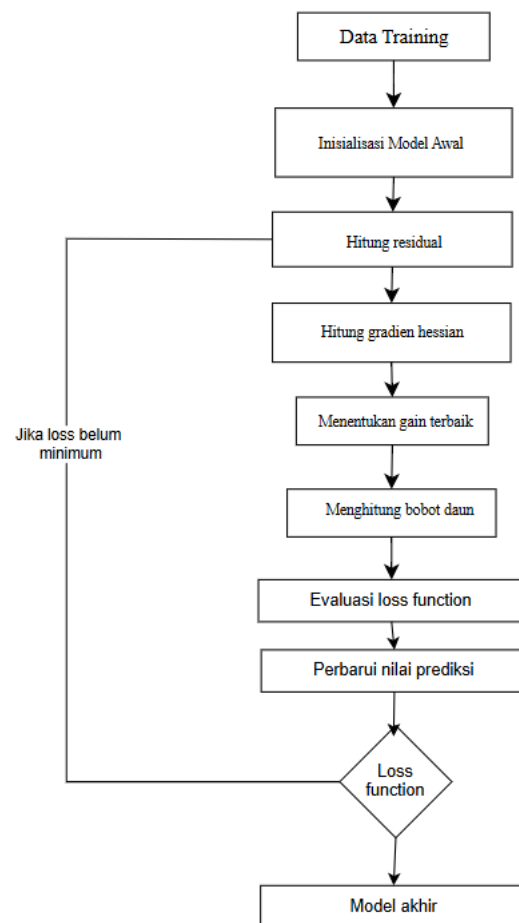
Setelah data melalui tahap preprocessing yang mencakup *data cleaning*, *data aggregation*, dan *data transformation*, selanjutnya dilakukan *feature engineering* dengan menambahkan beberapa fitur baru agar model dapat mengenali pola kunjungan dengan lebih baik. Fitur-fitur tersebut meliputi:

- Ekstraksi waktu, yaitu hari, bulan, tahun, *day_of_week*, dan *is_weekend*.
- Lag feature*, yaitu jumlah pengunjung pada 1 hari sebelumnya (*lag_1*) dan 7 hari sebelumnya (*lag_7*).
- Rolling statistic*, yaitu rata-rata jumlah pengunjung pada 7 hari terakhir (*rolling_7*) dan 30 hari terakhir (*rolling_30*).

Setelah seluruh tahapan *feature engineering* dilakukan, variabel-variabel hasil pengolahan data digunakan sebagai input (x), sedangkan jumlah pengunjung harian digunakan sebagai target (y). Data tersebut selanjutnya digunakan dalam

proses pelatihan dan pengujian model LightGBM, dengan skema pembagian data dan pengujian yang akan dijelaskan secara rinci pada subbab skenario pengujian.

Berikut adalah garis besar alur kerja dari algoritma LightGBM yang digunakan dalam penelitian ini :



Gambar 3.4 Alur Kerja *LightGBM*

Sebagai contoh, penulis akan melampirkan 12 sampel data hasil dari *feature engineering* yang akan digunakan sebagai input model *LightGBM*

Tabel 3.8 Sampel Data untuk simulasi perhitungan model LightGBM

Tanggal	Total Pengunjung	Waktu					Shock Feature	Lag feature		Rolling Statistic	
		Hari	Bulan	Tahun	<i>day_of_week</i>	<i>is_weekend</i>	<i>diff_pengunjung</i>	<i>lag_1</i>	<i>lag_7</i>	<i>rolling_7</i>	<i>rolling_30</i>
2/5/2023	25	Minggu	Februari	2023	6	1	-25	...	...	...	...
2/6/2023	0	Senin	Februari	2023	0	0	-13	25	...	...	...
2/7/2023	13	Selasa	Februari	2023	1	0	-11	0	...	...	...
2/8/2023	2	Rabu	Februari	2023	2	0	-2	13	...	...	...
2/9/2023	0	Kamis	Februari	2023	3	0	1	2	...	...	...
2/10/2023	1	Jum'at	Februari	2023	4	0	50	0	...	...	...
2/11/2023	51	Sabtu	Februari	2023	5	1	-43	1	...	...	...
2/12/2023	18	Minggu	Februari	2023	6	1	-18	51	25	15.57	...
2/13/2023	0	Senin	Februari	2023	0	0	1	18	0	12.14	...
2/14/2023	1	Selasa	Februari	2023	1	0	35	0	13	12.14	...
2/15/2023	36	Rabu	Februari	2023	2	0	-36	1	2	11.86	...
2/16/2023	0	Kamis	Februari	2023	3	0		36	0	15.29	...

Dari 12 sampel data di atas. Data yang digunakan sebagai contoh perhitungan model LightGBM adalah data pada 5 data terakhir yaitu, mulai tanggal 12 Februari 2023 sampai 16 Februari 2023 dengan hanya menggunakan 3 fitur yaitu *is_weekend*, *lag_1*, dan *rolling_7* sebagai gambaran dari perhitungan model LightGBM. Hal ini dikarenakan pada tahap contoh perhitungan, fitur *lag_7* dan *rolling_30* tidak digunakan terlebih dahulu agar proses penjelasan lebih sederhana dan mudah dipahami. Namun demikian, kedua fitur tersebut tetap akan dimanfaatkan pada tahap implementasi model secara menyeluruh di Bab 4, karena fitur *lag_7* mampu merepresentasikan pola kunjungan mingguan, sedangkan *rolling_30* membantu menangkap tren jangka panjang atau bulanan. Dengan demikian, penggunaan keduanya diharapkan dapat meningkatkan akurasi prediksi model.

Perhitungan model dimulai dengan menentukan y target dan $\hat{y}$ prediksi :

- a. Target y = jumlah pengunjung asli

$$y = [18,0,1,36,0]$$

- b. Prediksi awal $\hat{y}$ = rata-rata target :

$$\hat{y} = \frac{18 + 0 + 1 + 36 + 0}{5} = \frac{55}{5} = 11$$

Berdasarkan perhitungan di atas, prediksi awal untuk $\hat{y}$ di semua data adalah 11.

Selanjutnya dihitung nilai gradien (g_i) dan Hessian (h_i) berdasarkan fungsi objektif regresi Mean Squared Error (MSE). Pada objective regression, gradien dihitung menggunakan:

$$r = \hat{y}_i - y, h_i = 1 \quad (3.9)$$

Tabel hasil perhitungan gradien adalah sebagai berikut :

Tabel 3.9 Tabel Data gradien

Tanggal	y	$\hat{y}$	<i>residual</i>
2/12/2023	18	11	-7
2/13/2023	0	11	11
2/14/2023	1	11	10
2/15/2023	36	11	-25
2/16/2023	0	11	11

Selanjutnya, dicari split terbaik dengan menghitung nilai gain menggunakan rumus:

$$Gain = \frac{G_L^2}{H_L} + \frac{G_R^2}{H_R} - \frac{(G_L + G_R)^2}{H_L + H_R} \quad (3.10)$$

Fitur *is_weekend* terdiri dari 2 leaf, yaitu weekday (bernilai 0) dan weekend (bernilai 1). Perhitungan diperoleh sebagai berikut:

- a. Untuk weekday (13–16 Februari 2023) = [11, 10, (-25), 11]

$$G_{weekday} = 11 + 10 + (-25) + 11 = 7, H_{weekday} = 4$$

- b. Untuk weekend (12 Februari 2023): [(-7)]

$$G_{weekend} = -7, H_{weekend} = 1$$

Sehingga, gain untuk fitur *is_weekend* adalah:

$$Gain = \frac{7^2}{4} + \frac{(-7)^2}{1} - \frac{(7+(-7))^2}{4+1} = \frac{49}{4} + \frac{49}{1} + 0 = 12.25 + 49 = 61.25$$

Sementara itu, untuk fitur *lag_1* juga terbagi menjadi 2 leaf yaitu ≤ 5 dan >5 . Perhitungan diperoleh sebagai berikut :

- a. Untuk *lag_1* ≤ 5 (14 Februari 2023 dan 15 Februari 2023) = [10, (-25)]

$$G_{lag_{-1} \leq 5} = 10 + (-25) = -15, H_{lag_{-1} \leq 5} = 2$$

- b. Untuk *lag_1* > 5 (12 Februari 2023, 13 Februari 2023, dan 16 Februari 2023) =

$$[(-7), 11, 11]$$

$$G_{lag_{-1} > 5} = (-7) + 11 + 11 = 15, H_{lag_{-1} > 5} = 3$$

Sehingga, gain untuk fitur lag_1 adalah:

$$Gain = \frac{15^2}{3} + \frac{(-15)^2}{2} - \frac{(15+(-15))^2}{3+2} = \frac{225}{3} + \frac{225}{2} + 0 = 75 + 112.5 = 187.5$$

Sementara itu, untuk fitur $rolling_7$ juga terbagi menjadi 2 leaf yaitu ≤ 5 dan >5 . Perhitungan diperoleh sebagai berikut :

a. Untuk $rolling_7 \leq 15$ (13 Februari 2023 sampai 15 Februari 2023)

$$= [11, 10, (-25)]$$

$$G_{rolling_7 \leq 15} = 11 + 10 + (-25) = -4, H_{rolling_7 \leq 15} = 3$$

b. Untuk $rolling_7 > 15$ (12 Februari 2023 dan 16 Februari 2023) = $[(-7),$

$$11]$$

$$G_{rolling_7 > 15} = (-7) + 11 = 4, H_{rolling_7 > 15} = 2$$

Sehingga, gain untuk fitur $rolling_7$ adalah:

$$Gain = \frac{4^2}{3} + \frac{4^2}{2} - \frac{(4+(-4))^2}{3+2} = \frac{16}{3} + \frac{16}{2} + 0 = 5.33 + 8 = 13.33$$

Dari hasil perhitungan split ketiga fitur, dapat disimpulkan bahwa lag_1 memiliki nilai gain terbesar yaitu 187.5. Oleh karena itu, langkah selanjutnya adalah menghitung bobot daun optimal dengan menggunakan rumus:

$$w = -\frac{G}{H} \quad (3.11)$$

Berdasarkan hasil tersebut, fitur dengan nilai gain terbesar adalah lag_1 . Oleh karena itu, perhitungan bobot daun optimal dilakukan pada fitur ini dengan menggunakan rumus:

- a. *Left leaf* ($lag_1 \leq 5$)

$$w_{left} = - \frac{(-15)}{2} = 7.5$$

- b. *Right leaf* ($lag_1 > 5$)

$$w_{right} = - \frac{(15)}{3} = -5$$

Setelah bobot daun diperoleh, prediksi baru untuk setiap data dihitung kembali menggunakan rumus:

$$\hat{y}_{baru} = \hat{y}_{lama} + \eta \times w \quad (3.12)$$

Dengan nilai learning rate (η) = 0.1, diperoleh hasil prediksi untuk setiap data sebagai berikut:

- a. 14 Februari 2023 dan 15 Februari 2023 (*Right leaf*)

$$\hat{y}_{baru} = 11 + 0.1 \times (-5) = 10.5$$

- b. 12 Februari 2023, 13 Februari 2023, dan 16 Februari 2023 (*Left leaf*)

$$\hat{y}_{baru} = 11 + 0.1 \times 7.5 = 11.75$$

Sehingga, diperoleh ringkasan tabel dengan jumlah prediksi terbaru adalah:

Tabel 3.10 Tabel Data Jumlah Prediksi terbaru

Tanggal	y	$\hat{y}$	<i>residual</i>	Leaf	Bobot	Prediksi baru
2/12/2023	18	11	7	Left	7.5	11.75
2/13/2023	0	11	-11	Left	7.5	11.75
2/14/2023	1	11	-10	Right	-5	10.5
2/15/2023	36	11	25	Left	7.5	11.75
2/16/2023	0	11	-11	Right	-5	10.5

Secara konseptual, proses pembentukan model LightGBM dapat dilanjutkan pada pohon ke-2 dan seterusnya, di mana setiap pohon bertujuan untuk memperbaiki kesalahan prediksi dari pohon sebelumnya. Pada contoh perhitungan manual ini, proses hanya ditunjukkan hingga pohon pertama untuk mempermudah

pemahaman mekanisme kerja algoritma. Pada tahap implementasi menggunakan library LightGBM, jumlah iterasi atau pohon ditentukan melalui pengaturan hyperparameter tertentu, seperti jumlah pohon dan laju pembelajaran, serta dievaluasi berdasarkan performa model.

Untuk pohon pertama, perhitungan nilai MSE adalah sebagai berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.13)$$

Dengan data:

- a. Target y = jumlah pengunjung asli

$$y = [18, 0, 1, 36, 0]$$

- b. Prediksi $\hat{y}$ baru

$$\hat{y} = [11.75, 11.75, 10.5, 11.75, 10.5]$$

Maka diperoleh:

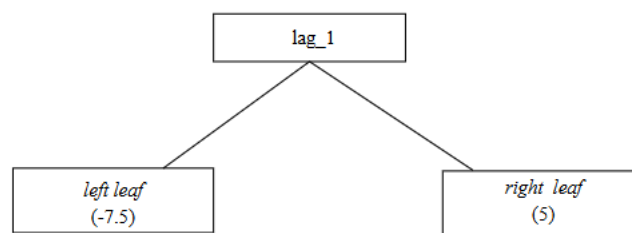
1. $(18 - 11.75)^2 = 39.06$
2. $(0 - 11.75)^2 = 138.06$
3. $(1 - 10.5)^2 = 90.25$
4. $(36 - 11.75)^2 = 588.06$
5. $(0 - 10.5)^2 = 110.25$

Sehingga :

$$MSE = \frac{39.06 + 138.06 + 90.25 + 588.06 + 110.25}{5} = 193.14$$

Berdasarkan perhitungan di atas, pada pohon pertama diperoleh nilai *Mean Squared Error* (MSE) sebesar 193.14. Nilai MSE tersebut ditampilkan untuk memberikan gambaran konseptual mengenai perhitungan error pada proses

pembentukan model LightGBM. Pada tahap implementasi menggunakan *library* LightGBM, model dilatih dengan konfigurasi hyperparameter tertentu, dan evaluasi performa dilakukan berdasarkan hasil prediksi pada data uji untuk setiap skenario pengujian.



Gambar 3.5 Ilustrasi Pohon Keputusan Pertama

Gambar di atas memperlihatkan bagaimana algoritma LightGBM memilih fitur dengan nilai gain tertinggi, yaitu *lag_1*, sebagai node pertama (root). Selanjutnya, fitur *rolling_7* dan *is_weekend* digunakan pada node berikutnya untuk membentuk cabang hingga diperoleh empat leaf dengan nilai prediksi akhir yang berbeda pada setiap jalur pohon.

3.2.5 Skenario Pengujian

Skenario pengujian disusun untuk mengevaluasi kinerja model LightGBM dalam memprediksi jumlah pengunjung museum berdasarkan data historis yang telah melalui tahap preprocessing dan feature engineering. Pengujian dilakukan untuk mengetahui kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya serta untuk membandingkan performa model pada beberapa skema pembagian data.

Pada tahap implementasi, proses pelatihan dan pengujian model dilakukan menggunakan *library* LightGBM dengan konfigurasi hyperparameter tertentu. Evaluasi performa model dilakukan berdasarkan hasil prediksi pada data uji menggunakan metrik evaluasi yang ditetapkan. Untuk memperoleh hasil yang optimal dan objektif, penelitian ini menerapkan beberapa skenario pengujian yang dibedakan berdasarkan proporsi data latih dan data uji.

Rincian skenario pengujian yang digunakan dalam penelitian ini, yang dibedakan berdasarkan proporsi data training dan data testing, disajikan pada Tabel 3.11.

Tabel 3.11 Tabel Skenario Pengujian

Skenario	Data Training	Data Testing
Skenario A	70%	30%
Skenario B	80%	20%
Skenario C	90%	10%

Ketiga skenario tersebut digunakan untuk menganalisis pengaruh perbedaan proporsi data training dan testing terhadap tingkat kesalahan prediksi dan stabilitas kinerja model LightGBM.

Selain variasi pembagian data, kinerja model LightGBM juga dipengaruhi oleh pengaturan hyperparameter yang digunakan selama proses pelatihan. Oleh karena itu, pada setiap skenario pengujian dilakukan proses tuning hyperparameter untuk memperoleh konfigurasi model yang optimal dan menghasilkan performa prediksi yang lebih baik (Wang & Wang, 2020).

Tuning hyperparameter dilakukan untuk meningkatkan performa model LightGBM dengan cara mencoba beberapa kombinasi nilai hyperparameter tertentu. Hyperparameter merupakan parameter yang ditentukan sebelum proses

pelatihan model dan berpengaruh terhadap kompleksitas serta kemampuan generalisasi model (Aulady, 2024).

Dalam penelitian ini, tuning hyperparameter dilakukan pada setiap skenario pengujian. Proses tuning bertujuan untuk memperoleh kombinasi hyperparameter yang menghasilkan nilai kesalahan prediksi paling kecil berdasarkan metrik evaluasi Mean Squared Error (MSE).

Tabel 3.12 Tabel Hyperparameter

Hyperparameter	Nilai Hyperparameter	Fungsi
<i>learning_rate</i>	0.01, 0.05, 0.1	Mengontrol laju pembelajaran model
<i>n_estimators</i>	100, 200, 300	Jumlah pohon dalam model
<i>max_depth</i>	3, 5, 7	Kedalaman maksimum setiap pohon
<i>num_leaves</i>	15, 31, 63	Jumlah maksimum daun pada pohon
<i>min_data_in_leaf</i>	20, 30, 50	Jumlah minimum data pada setiap daun

Kombinasi hyperparameter tersebut diuji pada setiap skenario pengujian untuk membandingkan performa model LightGBM pada berbagai kondisi pengujian. Melalui perbandingan ini, dapat dianalisis pengaruh perbedaan skenario dan konfigurasi parameter terhadap hasil prediksi, sehingga dapat ditentukan skenario serta konfigurasi model yang menghasilkan kinerja prediksi paling optimal (Bombrun et al., 2020).

3.2.6 Evaluasi Model

Pada tahap ini dilakukan evaluasi performa model LightGBM dalam memprediksi jumlah pengunjung harian Museum Mpu Purwa. Evaluasi dilakukan terhadap hasil prediksi pada data uji menggunakan model dengan konfigurasi terbaik yang diperoleh dari skenario pengujian sebelumnya. Evaluasi ini bertujuan

untuk menilai sejauh mana model mampu menghasilkan prediksi yang akurat dan stabil terhadap data yang belum pernah digunakan dalam proses pelatihan.

Evaluasi model pada penelitian ini menggunakan tiga metrik utama, yaitu Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan Mean Absolute Percentage Error (MAPE). Ketiga metrik ini dipilih karena mampu menggambarkan kinerja model dari sisi rata-rata kesalahan, sensitivitas terhadap error bernilai besar, serta kesalahan dalam bentuk persentase yang lebih mudah diinterpretasikan.

a. *Mean Absolute Error (MAE)*

MAE adalah metrik yang menghitung rata-rata selisih absolut antara nilai aktual dengan nilai prediksi. Metrik ini menggambarkan seberapa besar kesalahan rata-rata model dalam melakukan prediksi, tanpa memperhatikan arah kesalahan (positif atau negatif).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.14)$$

Semakin kecil nilai MAE, semakin baik performa model.

b. *Root Mean Squared Error (RMSE)*

RMSE merupakan akar dari rata-rata kuadrat selisih antara nilai aktual dengan nilai prediksi. RMSE lebih sensitif terhadap kesalahan besar karena selisih dikuadratkan, sehingga error yang besar akan berpengaruh signifikan pada hasil akhir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.15)$$

Semakin kecil nilai RMSE, semakin akurat model dalam melakukan prediksi.

c. *Mean Absolute Percentage Error (MAPE)*

MAPE adalah metrik yang mengukur rata-rata kesalahan prediksi dalam bentuk persentase terhadap nilai aktual. Dengan demikian, MAPE lebih mudah diinterpretasikan karena hasilnya dalam skala persen.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3.16)$$

Semakin kecil nilai MAPE, semakin akurat prediksi yang dihasilkan model.

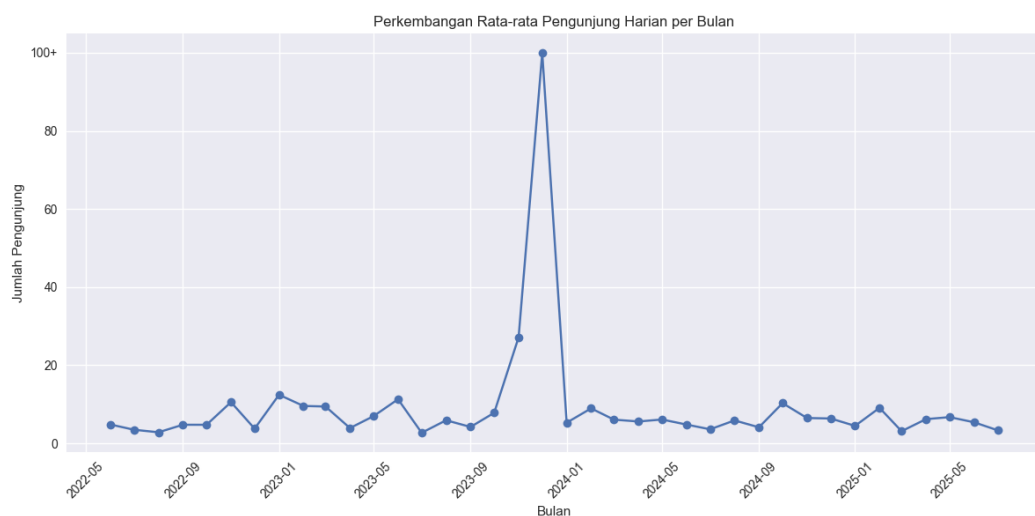
Dengan menggunakan ketiga metrik evaluasi tersebut, performa model dapat dinilai dari berbagai sisi, baik rata-rata kesalahan, sensitivitas terhadap error besar, maupun kesalahan dalam bentuk persentase.

BAB IV

HASIL DAN PEMBAHASAN

4.1 *Exploratory Data Analysis (EDA)*

Dataset yang digunakan dalam penelitian ini merupakan data jumlah pengunjung harian Museum Mpu Purwa dengan periode pengamatan tahun 2022 hingga 2025. Setelah proses seleksi atribut, data yang digunakan berjumlah sebanyak 6.161 data dengan dua atribut utama, yaitu *timestamp* dan *jumlah_pengunjung*. Data kemudian diagregasi menjadi total jumlah pengunjung per hari sehingga setiap baris merepresentasikan satu tanggal observasi. Tahap ini bertujuan untuk memastikan kualitas data, mengetahui pola distribusi.



Gambar 4.1 Grafik Rata-Rata Jumlah Pengunjung

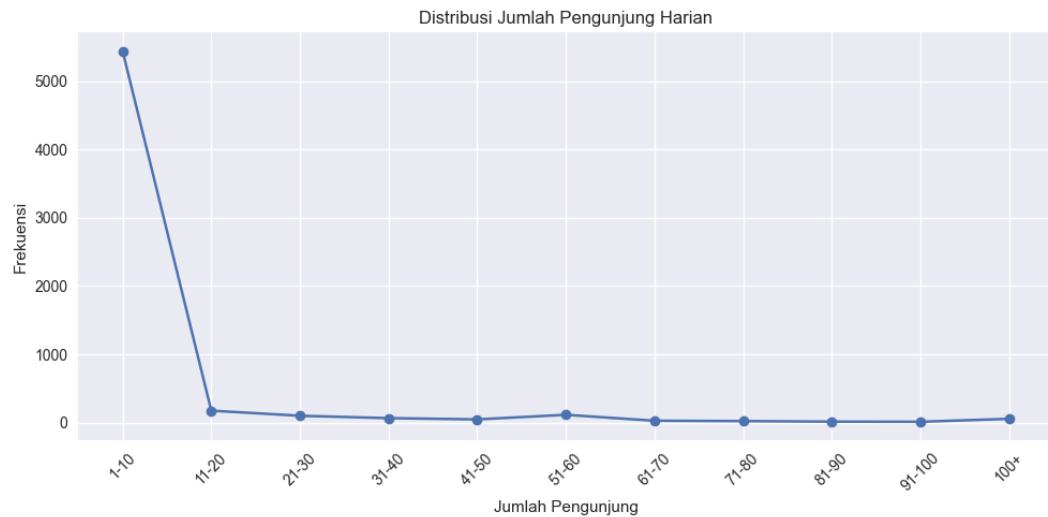
Berdasarkan gambar 4.1, terlihat bahwa jumlah pengunjung museum cenderung berada pada kisaran yang relatif rendah dan stabil, yaitu sekitar 3 hingga 12 pengunjung per hari pada sebagian besar periode pengamatan. Namun,

terdapat satu lonjakan yang sangat signifikan pada akhir tahun 2023, di mana rata-rata pengunjung meningkat drastis hingga mencapai batas visualisasi yaitu lebih dari 100 kunjungan. Lonjakan ini kemungkinan dipengaruhi oleh faktor tertentu, seperti adanya event khusus, masa liburan, atau anomali data yang belum sepenuhnya teridentifikasi.

Setelah periode lonjakan tersebut, rata-rata jumlah pengunjung kembali berada pada kisaran normal di bawah 10 pengunjung per hari. Secara umum, grafik menunjukkan bahwa tingkat kunjungan museum tidak mengalami peningkatan yang konsisten, melainkan berfluktuasi dalam rentang yang relatif kecil. Pola ini mengindikasikan bahwa kunjungan museum masih bersifat sporadis dan belum menunjukkan tren pertumbuhan yang signifikan dalam jangka panjang.

Setelah menganalisis perkembangan rata-rata jumlah pengunjung museum dari waktu ke waktu, langkah selanjutnya adalah mengamati pola persebaran data melalui analisis distribusi jumlah pengunjung harian. Analisis distribusi ini bertujuan untuk mengetahui rentang jumlah pengunjung yang paling sering terjadi serta melihat apakah data cenderung terkonsentrasi pada nilai tertentu atau tersebar

secara merata. Dengan memahami distribusi data, dapat diperoleh gambaran yang lebih jelas mengenai karakteristik kunjungan harian di Museum Mpu Purwa



Gambar 4.2 Distribusi Jumlah Pengunjung

Berdasarkan grafik distribusi jumlah pengunjung harian, terlihat bahwa sebagian besar data terkonsentrasi pada kategori 1–10 pengunjung per hari, dengan frekuensi yang jauh lebih tinggi dibandingkan kategori lainnya. Hal ini menunjukkan bahwa pada mayoritas hari, jumlah pengunjung museum berada dalam kisaran yang relatif rendah.

Frekuensi kemudian menurun secara signifikan pada kategori 11–20 dan seterusnya, serta semakin kecil pada rentang yang lebih tinggi. Meskipun terdapat beberapa hari dengan jumlah pengunjung yang lebih besar, termasuk kategori 100 ke atas kunjungan, jumlah tersebut relatif sedikit dibandingkan keseluruhan data. Pola ini menunjukkan bahwa distribusi data bersifat right-skewed (miring ke

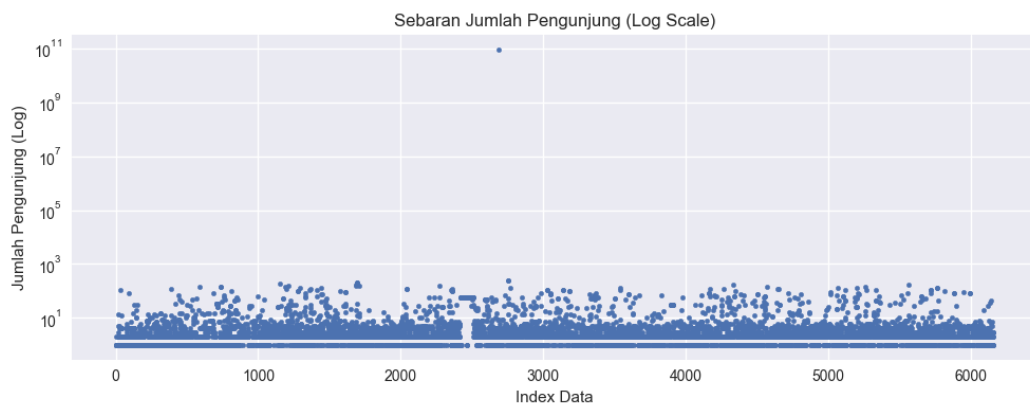
kanan), di mana sebagian besar nilai berada pada rentang rendah dengan sedikit nilai ekstrem pada rentang tinggi.

Secara keseluruhan, distribusi ini mengindikasikan bahwa kunjungan harian Museum Mpu Purwa didominasi oleh jumlah pengunjung yang rendah, sementara lonjakan pengunjung dalam jumlah besar hanya terjadi pada kondisi tertentu dan tidak bersifat rutin.

Setelah menganalisis distribusi jumlah pengunjung harian, tahap selanjutnya adalah mengidentifikasi kemungkinan adanya nilai ekstrem (*outlier*) dalam dataset. Keberadaan outlier penting untuk diperhatikan karena dapat memengaruhi nilai rata-rata, standar deviasi, serta menghasilkan interpretasi yang kurang representatif terhadap kondisi sebenarnya. Oleh karena itu, diperlukan analisis lanjutan untuk memastikan apakah terdapat data yang menyimpang secara signifikan dari pola umum.

Untuk memperoleh gambaran yang lebih jelas mengenai pola sebaran data serta mendeteksi nilai yang menyimpang secara ekstrem, digunakan visualisasi *scatter plot* dengan skala logaritmik pada sumbu vertikal. Penggunaan skala logaritmik bertujuan untuk mengurangi distorsi akibat perbedaan skala yang sangat

besar antar nilai, sehingga mayoritas data tetap dapat diamati dengan jelas meskipun terdapat nilai yang jauh lebih besar dibandingkan yang lain.



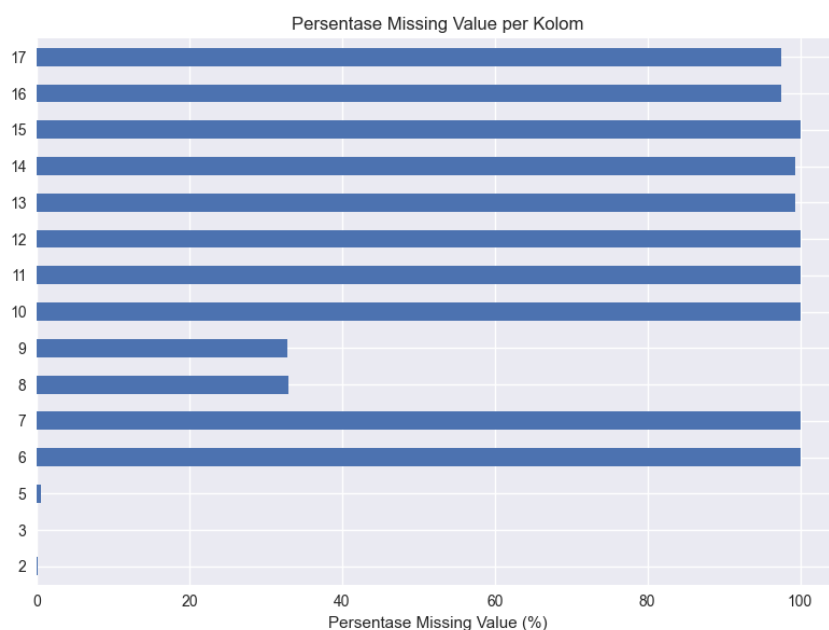
Gambar 4.3 Deteksi Outlier Jumlah Pengunjung

Berdasarkan gambar 4.3, terlihat bahwa sebagian besar data terkonsentrasi pada rentang nilai yang relatif rendah, yaitu di bawah 100 pengunjung per hari. Titik-titik data tampak menyebar secara padat di bagian bawah grafik, menunjukkan bahwa mayoritas hari memiliki jumlah pengunjung yang kecil.

Namun demikian, terdapat satu titik yang berada sangat jauh di atas kumpulan data lainnya, yang mengindikasikan adanya nilai ekstrem (outlier). Perbedaan yang sangat signifikan antara nilai tersebut dan mayoritas data menunjukkan bahwa angka tersebut tidak mengikuti pola umum kunjungan harian museum. Oleh karena itu, nilai tersebut perlu ditinjau lebih lanjut untuk memastikan apakah merupakan kejadian khusus atau kesalahan pencatatan data.

Penggunaan skala logaritmik pada sumbu vertikal membantu memperjelas perbedaan skala antara data normal dan nilai ekstrem, sehingga pola distribusi mayoritas data tetap dapat diamati tanpa terdistorsi oleh outlier.

Setelah mengidentifikasi keberadaan outlier dalam dataset, tahap selanjutnya adalah menganalisis kualitas data melalui pemeriksaan missing value pada setiap kolom. Analisis ini bertujuan untuk mengetahui apakah terdapat data yang tidak terisi, serta untuk menentukan langkah penanganan yang tepat sebelum dilakukan analisis lanjutan.



Gambar 4.4 Presentase Missing Value per Kolom

Berdasarkan grafik persentase missing value per kolom, terlihat bahwa sebagian besar kolom memiliki tingkat missing value yang sangat tinggi, bahkan mendekati atau mencapai 100%. Hal ini menunjukkan bahwa kolom-kolom tersebut hampir tidak memiliki data yang terisi sehingga kurang relevan untuk digunakan dalam proses analisis.

Selain itu, terdapat beberapa kolom dengan persentase missing pada kisaran menengah (sekitar 30–40%), yang masih memerlukan pertimbangan lebih lanjut apakah akan dilakukan imputasi atau penghapusan. Sementara itu, hanya sedikit

kolom yang memiliki tingkat missing value sangat rendah atau mendekati 0%, yang menunjukkan bahwa kolom tersebut memiliki kualitas data yang relatif baik dan lebih layak digunakan dalam analisis.

Secara keseluruhan, tingginya persentase missing value pada sebagian besar kolom mengindikasikan bahwa proses pembersihan data (data cleaning) menjadi tahap yang krusial sebelum dilakukan analisis lanjutan, terutama dalam menentukan kolom mana yang relevan dan memiliki kelengkapan data yang memadai.

Secara umum, hasil *Exploratory Data Analysis* (EDA) menunjukkan bahwa pola jumlah pengunjung museum cenderung berada pada rentang rendah dengan fluktuasi yang relatif kecil dari waktu ke waktu. Distribusi data bersifat *right-skewed*, di mana sebagian besar nilai terkonsentrasi pada jumlah pengunjung yang rendah, sementara terdapat beberapa nilai ekstrem yang menyimpang jauh dari pola mayoritas data. Keberadaan nilai ekstrem tersebut berpotensi memengaruhi perhitungan statistik seperti rata-rata dan standar deviasi.

Selain itu, hasil analisis missing value menunjukkan bahwa sebagian besar atribut dalam dataset memiliki persentase nilai kosong yang sangat tinggi, bahkan mendekati 100% pada beberapa kolom. Kondisi ini mengindikasikan bahwa tidak seluruh atribut layak digunakan dalam proses analisis dan pemodelan.

Oleh karena itu, diperlukan tahap *data preprocessing* yang meliputi pembersihan data, penanganan outlier, serta seleksi atau penghapusan atribut dengan tingkat missing value yang tinggi. Tahap ini bertujuan untuk memastikan bahwa data yang digunakan dalam proses pembentukan model telah bersih, relevan,

dan representatif sehingga dapat menghasilkan performa model yang lebih optimal dan akurat.

4.2 Hasil *Preprocessing Data*

Tahap preprocessing dilakukan untuk menyiapkan data sebelum proses pemodelan. Proses ini meliputi *data cleaning*, *aggregation data*, serta penerapan *capping* untuk mengatasi nilai ekstrem agar data lebih siap dan stabil digunakan dalam pembentukan model.

4.2.1 Hasil *Cleaning Data*

Tahap data cleaning dilakukan untuk memastikan dataset memenuhi standar kualitas dan siap digunakan dalam proses pemodelan. Proses ini mencakup penghapusan atribut dengan persentase missing value yang tinggi, penanganan outlier, serta seleksi fitur berdasarkan relevansi terhadap tujuan penelitian. Berdasarkan hasil *Exploratory Data Analysis* (EDA), sebagian besar atribut awal tidak memiliki kelengkapan data yang memadai dan tidak berkaitan langsung dengan variabel yang diprediksi. Oleh karena itu, dipilih dua atribut utama yang paling relevan, yaitu *timestamp* yang setelah preprocessing diubah menjadi tanggal sebagai variabel waktu, serta *jumlah_pengunjung* sebagai variabel target. Selain itu, dilakukan pemeriksaan untuk memastikan tidak terdapat missing value pada kedua atribut tersebut.

Tabel 4.1 Data setelah *cleaning*

tanggal	jumlah_pengunjung
2022-07-01	21
2022-07-02	22
2022-07-03	8
2022-07-04	4

tanggal	jumlah_pengunjung
2022-07-05	12
2022-07-06	23
2022-07-07	95
2022-07-08	8
2022-07-12	10
2022-07-13	15
2022-07-14	1
....	
2025-07-20	37

Pada Tabel 4.1 ditampilkan hasil dataset setelah melalui tahap *data cleaning*, yang menunjukkan bahwa data akhir hanya terdiri dari dua atribut utama, yaitu tanggal sebagai penanda waktu kunjungan dan jumlah_pengunjung sebagai representasi jumlah pengunjung pada waktu tertentu. Berdasarkan hasil pemeriksaan, tidak ditemukan missing value pada kedua atribut tersebut sehingga tidak diperlukan proses imputasi, dan dataset dinyatakan siap untuk digunakan pada tahap transformasi data serta pembangunan model prediksi.

4.2.2 Hasil Agregation Data

Tahap aggregation data dilakukan untuk mengubah data mentah berbasis waktu (*timestamp*) menjadi data harian yang lebih terstruktur. Proses ini dilakukan dengan mengonversi atribut *timestamp* menjadi format tanggal (tanpa informasi jam), kemudian melakukan pengelompokan berdasarkan tanggal serta menjumlahkan nilai jumlah_pengunjung pada setiap tanggal yang sama. Hasil dari proses ini menghasilkan data harian yang lebih ringkas dan sesuai untuk analisis deret waktu (*time series*).

Tabel 4.2 Data setelah *Agregation*

tanggal	jumlah_pengunjung
2022-08-02	5
2022-08-03	10
2022-08-04	5
2022-08-05	9
2022-08-06	8
2022-08-07	14
2022-08-09	64
2022-08-10	18
2022-08-11	2
2022-08-12	1
2022-08-13	21
2022-08-14	10
2022-08-16	27
2022-08-18	16
2022-08-19	4
....	
2025-07-20	37

Pada Tabel 4.2 ditampilkan hasil dataset setelah proses agregasi berdasarkan tanggal, di mana setiap baris merepresentasikan total jumlah pengunjung dalam satu hari sehingga data menjadi lebih ringkas dan terstruktur. Jumlah data yang semula terdiri dari ribuan baris kejadian individu berkurang menjadi 875 baris data harian, sehingga memudahkan analisis pola kunjungan seperti tren, musiman, dan fluktuasi harian, serta lebih sesuai digunakan dalam tahap pemodelan prediksi berbasis *time series*.

4.2.3 Hasil *Capping Data*

Tahap *capping* dilakukan sebagai bagian dari preprocessing untuk menangani nilai ekstrem (*outlier*) pada variabel jumlah_pengunjung yang teridentifikasi melalui *Exploratory Data Analysis* (EDA). Beberapa nilai yang sangat tinggi dan tidak proporsional berpotensi menyebabkan bias pada model prediksi, sehingga diterapkan teknik *capping* dengan membatasi nilai yang melebihi ambang batas tertentu menjadi nilai maksimum yang telah ditentukan. Penentuan

batas dilakukan menggunakan pendekatan statistik, sehingga nilai ekstrem tidak dihapus, melainkan disesuaikan agar distribusi data tetap representatif tanpa kehilangan informasi penting.

Tabel 4.3 Data Setelah *Capping*

tanggal	jumlah_pengunjung	Jumlah_pengunjung_capped
2022-06-17	14	14
2022-06-18	2	2
2022-06-19	3	3
2022-06-21	15	15
2022-06-23	4	4
2022-06-24	8	8
2022-06-25	15	15
2022-06-26	106	98.5
2022-06-28	4	4
2022-06-29	12	12
2022-06-30	14	14
2022-07-01	21	21
2022-07-02	22	22
2022-07-03	8	8
2022-07-04	4	4
...	...	...
2025-07-20	37	37

Pada Tabel 4.3 ditampilkan hasil data setelah proses capping, di mana sebagian besar nilai tidak mengalami perubahan karena masih berada dalam rentang normal, sedangkan nilai ekstrem seperti pada tanggal 2022-06-26 disesuaikan dari 106 menjadi 98.5 sesuai batas maksimum yang ditentukan. Setelah proses ini, distribusi data menjadi lebih stabil dan tidak lagi didominasi oleh nilai ekstrem, sehingga dataset lebih siap digunakan dalam tahap pemodelan deret waktu untuk menghasilkan prediksi yang lebih akurat dan robust terhadap anomali data.

4.3 *Feature Engineering*

Tahap *feature engineering* dilakukan untuk menambahkan informasi baru ke dalam dataset sebelum proses pemodelan. Fitur yang dibentuk meliputi ekstraksi

fitur waktu dari atribut *timestamp*, pembuatan *shock feature*, pembentukan *lag feature*, serta perhitungan *rolling mean* berdasarkan jumlah pengunjung hariannya.

4.3.1 Hasil Ekstrasi Fitur Waktu

Pada tahap ekstrasi fitur waktu, kolom tanggal diolah untuk menghasilkan beberapa fitur tambahan, yaitu *day*, *month*, dan *year*, yang merepresentasikan hari, bulan, dan tahun terjadinya kunjungan. Selain itu, ditambahkan pula fitur *day_of_week* untuk menunjukkan hari dalam seminggu (0–6) serta fitur *is_weekend* sebagai indikator apakah suatu tanggal termasuk akhir pekan atau tidak. Penambahan fitur-fitur ini bertujuan untuk membantu model dalam mengidentifikasi pola kunjungan berdasarkan waktu, seperti perbedaan jumlah pengunjung pada hari kerja dan akhir pekan.

Tabel 4.4 Hasil Ekstrasi Fitur Waktu

tanggal	day	month	year	day_of_week	is_weekend
2022-07-24	24	7	2022	6	1
2022-07-26	26	7	2022	1	0
2022-07-27	27	7	2022	2	0
2022-07-28	28	7	2022	3	0
2022-07-29	29	7	2022	4	0
2022-07-31	31	7	2022	6	1
2022-08-02	2	8	2022	1	0
2022-08-03	3	8	2022	2	0
2022-08-04	4	8	2022	3	0
2022-08-05	5	8	2022	4	0
2022-08-06	6	8	2022	5	1
2022-08-07	7	8	2022	6	1
2022-08-09	9	8	2022	1	0
...	...	...	...	...	...
2025-07-20	20	7	2025	6	1

Berdasarkan Tabel 4.4, setiap data kini tidak hanya memuat informasi tanggal, tetapi juga rincian komponen waktu yang lebih spesifik seperti hari, bulan,

dan tahun. Fitur *day_of_week* memungkinkan model mengenali pola kunjungan berdasarkan hari tertentu, sedangkan fitur *is_weekend* membantu mengidentifikasi perbedaan pola antara hari kerja dan akhir pekan.

4.3.2 Hasil *Shock Feature*

Pada tahap ini dibentuk fitur *shock* yang merepresentasikan perubahan jumlah pengunjung dibandingkan hari sebelumnya. Fitur ini dihitung berdasarkan selisih nilai *jumlah_pengunjung_capped* antar dua hari yang berurutan.

Tabel 4.5 Hasil *Shock Feature*

tanggal	diff_pengunjung
2022-07-24	1
2022-07-26	3
2022-07-27	26
2022-07-28	4
2022-07-29	-25
2022-07-31	-5
2022-08-02	-2
2022-08-03	5
2022-08-04	-5
2022-08-05	4
2022-08-06	-1
2022-08-07	6
2022-08-09	50
...	...
2025-07-20	1

Berdasarkan Tabel 4.5, kolom *diff_pengunjung* menunjukkan besarnya perubahan jumlah pengunjung dari satu hari ke hari berikutnya. Nilai positif mengindikasikan adanya peningkatan jumlah pengunjung dibandingkan hari sebelumnya, sedangkan nilai negatif menunjukkan terjadinya penurunan.

Sebagai contoh, nilai 26 pada tanggal 2022-07-27 menunjukkan adanya kenaikan signifikan dibandingkan hari sebelumnya. Sebaliknya, nilai -25 pada tanggal 2022-07-29 menunjukkan penurunan jumlah pengunjung yang cukup besar.

Variasi nilai ini menggambarkan dinamika kunjungan museum yang fluktuatif dari hari ke hari.

4.3.3 Hasil *Lag Feature*

Pada tahap ini dibentuk fitur *lag* untuk merepresentasikan nilai jumlah pengunjung pada periode sebelumnya. Fitur yang digunakan terdiri dari *lag_1* dan *lag_7*, yang masing-masing menunjukkan jumlah pengunjung satu hari sebelumnya dan tujuh hari sebelumnya.

Tabel 4.6 Hasil *Lag Feature*

tanggal	lag_1	lag_7
2022-07-24	3	44
2022-07-26	4	27
2022-07-27	7	20
2022-07-28	33	34
2022-07-29	37	7
2022-07-31	12	46
2022-08-02	7	3
2022-08-03	5	4
2022-08-04	10	7
2022-08-05	5	33
2022-08-06	9	37
...	...	...
2025-07-20	98.5	11

Berdasarkan Tabel 4.6, kolom *lag_1* dan *lag_7* merepresentasikan nilai jumlah pengunjung pada periode sebelumnya yang digunakan sebagai referensi dalam memprediksi nilai pada hari berjalan. Sebagai contoh, nilai pada kolom *lag_1* menunjukkan jumlah pengunjung pada satu hari sebelum tanggal tersebut, sedangkan *lag_7* menunjukkan jumlah pengunjung pada tujuh hari sebelumnya.

Perbedaan nilai antara *lag_1* dan *lag_7* juga dapat menggambarkan adanya pola musiman mingguan. Jika nilai pada *lag_7* memiliki keterkaitan dengan nilai

saat ini, maka dapat diindikasikan adanya pola kunjungan yang berulang setiap minggu.

4.3.4 Hasil *Rolling Mean*

Pada tahap ini dibentuk fitur *rolling mean* dengan jendela 7 hari dan 30 hari berdasarkan nilai *jumlah_pengunjung_capped*. Kolom *rolling_mean_7* merepresentasikan rata-rata jumlah pengunjung dalam tujuh hari sebelumnya, sedangkan *rolling_mean_30* menunjukkan rata-rata dalam tiga puluh hari sebelumnya.

Tabel 4.7 Hasil *Rolling Mean*

tanggal	rolling mean 7	rolling mean 30
2022-07-24	20.14285714	19.78333333
2022-07-26	17.28571429	19.55
2022-07-27	19.14285714	20.58333333
2022-07-28	19.57142857	21.71666667
2022-07-29	20.28571429	21.61666667
2022-07-31	14.71428571	21.71666667
2022-08-02	15	21.61666667
2022-08-03	15.85714286	21.45
2022-08-04	15.57142857	18.33333333
2022-08-05	12.14285714	18.5
2022-08-06	8	18.36666667
...	...	...
2025-07-20	33.78571429	28.51667

Berdasarkan hasil pada Tabel 4.7, terlihat bahwa nilai *rolling_mean_7* cenderung lebih fluktuatif dibandingkan *rolling_mean_30*. Hal ini karena jendela 7 hari merepresentasikan pola jangka pendek yang lebih sensitif terhadap perubahan harian. Sebaliknya, *rolling_mean_30* menghasilkan nilai yang lebih stabil karena mencerminkan kecenderungan kunjungan dalam periode yang lebih panjang.

Sebagai contoh, pada tanggal 2022-07-24 nilai *rolling_mean_7* sebesar 20.14285714 dan *rolling_mean_30* sebesar 19.78333. Perbedaan ini menunjukkan bahwa rata-rata kunjungan dalam 7 hari terakhir sedikit lebih tinggi dibandingkan

rata-rata 30 hari terakhir. Pada beberapa tanggal lainnya terlihat adanya penurunan nilai *rolling_mean_7*, yang mengindikasikan melemahnya tren kunjungan dalam jangka pendek, meskipun nilai *rolling_mean_30* relatif lebih stabil.

4.4 Skenario Pengujian

Pada bagian ini dibahas hasil evaluasi performa model LightGBM berdasarkan beberapa skenario pembagian data latih dan data uji serta kombinasi hyperparameter yang digunakan. Hasil evaluasi diperoleh dari perbandingan antara nilai aktual (y_{test}) dan nilai prediksi (y_{pred}) untuk mengukur tingkat akurasi model dalam memprediksi jumlah pengunjung harian. Evaluasi performa model berfokus pada nilai RMSE sebagai indikator utama dalam menentukan model terbaik pada masing-masing skenario pengujian.

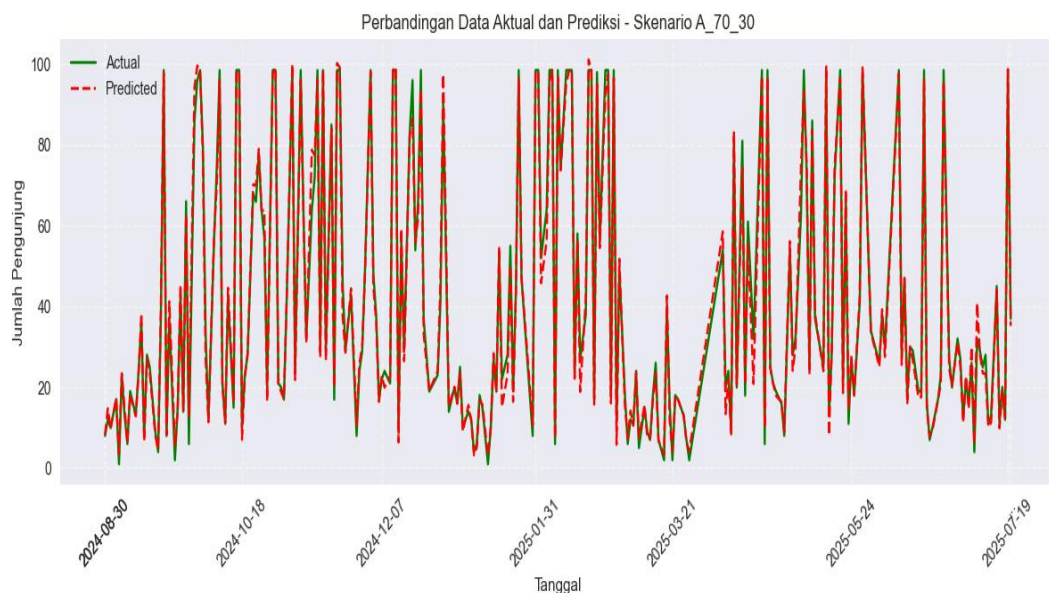
4.4.1 Skenario Pengujian A

Pada skenario pengujian A, data dibagi menjadi 70% sebagai data latih dan 30% sebagai data uji berdasarkan urutan waktu (*time-based split*). Proporsi ini memungkinkan model mempelajari pola historis dari sebagian besar data, kemudian diuji pada periode berikutnya untuk melihat kemampuan generalisasi dalam memprediksi jumlah pengunjung harian.

Berdasarkan hasil pengujian terhadap 243 kombinasi hyperparameter, diperoleh kombinasi terbaik dengan *learning_rate* sebesar 0.05, *n_estimators* sebanyak 300, *max_depth* sebesar 5, *num_leaves* sebanyak 31, dan *min_data_in_leaf* sebesar 20. Model tersebut menghasilkan nilai MAE sebesar 2,08, RMSE sebesar 3.03, dan MAPE sebesar 11.54%.

Nilai RMSE sebesar 3.03 menunjukkan bahwa rata-rata kesalahan prediksi model terhadap nilai aktual berada pada kisaran tiga pengunjung per hari. Sementara itu, nilai MAE sebesar 2.08 mengindikasikan bahwa secara rata-rata selisih absolut antara prediksi dan nilai aktual relatif kecil. Nilai MAPE sebesar 11.54% menunjukkan bahwa tingkat kesalahan prediksi masih berada dalam batas yang wajar untuk data dengan karakteristik fluktuatif seperti jumlah pengunjung harian.

Berdasarkan nilai evaluasi tersebut, diperlukan analisis visual untuk melihat bagaimana performa model dalam mengikuti pola pergerakan data aktual. Visualisasi perbandingan antara nilai aktual dan nilai prediksi digunakan untuk mengidentifikasi sejauh mana model mampu menangkap tren, fluktuasi, serta lonjakan jumlah pengunjung dari waktu ke waktu.



Gambar 4.5 Grafik Perbandingan Data Aktual dan Prediksi Skenario A

Berdasarkan Gambar 4.5, data aktual jumlah pengunjung harian menunjukkan pola yang sangat fluktuatif dengan beberapa lonjakan tinggi dan

penurunan yang tajam pada periode tertentu. Variasi nilai yang cukup ekstrem ini menunjukkan bahwa jumlah pengunjung tidak memiliki pola yang stabil, sehingga menjadi tantangan bagi model dalam melakukan prediksi secara akurat.

Pada Gambar 4.6 juga terlihat bahwa garis prediksi mampu mengikuti pola umum dari data aktual, terutama dalam menangkap tren naik dan turun jumlah pengunjung. Namun, nilai prediksi terlihat lebih halus dibandingkan data aktual, sehingga pada beberapa titik lonjakan yang sangat tinggi maupun penurunan ekstrem masih terdapat selisih antara nilai prediksi dan nilai sebenarnya. Meskipun demikian, pola fluktuasi yang dihasilkan model tetap sejalan dengan pergerakan data aktual secara keseluruhan.

Secara umum, kesesuaian pola antara data aktual dan data prediksi menunjukkan bahwa model mampu mempelajari karakteristik utama dari data historis. Hal ini juga didukung oleh nilai evaluasi RMSE, MAE, dan MAPE yang relatif kecil, sehingga dapat disimpulkan bahwa model pada skenario pembagian data 70:30 memiliki performa yang cukup baik dalam memprediksi jumlah pengunjung harian meskipun terdapat perbedaan pada beberapa titik ekstrem.

4.4.2 Skenario Pengujian B

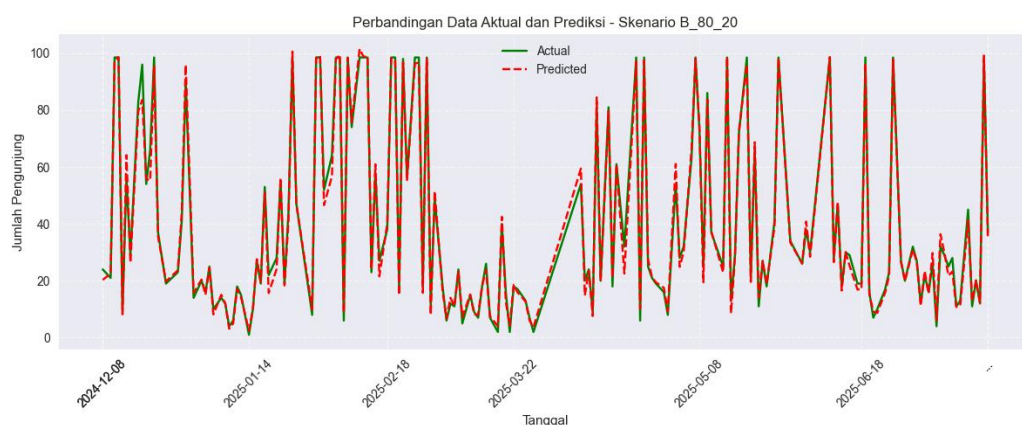
Pada skenario pengujian B, dataset dibagi menjadi 80% data latih dan 20% data uji berdasarkan urutan waktu. Dengan proporsi data latih yang lebih besar dibandingkan skenario A, model memiliki lebih banyak informasi historis untuk mempelajari pola pergerakan jumlah pengunjung sebelum dilakukan pengujian pada data periode berikutnya.

Hasil pengujian menunjukkan bahwa kombinasi hyperparameter terbaik pada skenario ini adalah *learning_rate* sebesar 0.01, *n_estimators* sebanyak 300,

max_depth sebesar 7, *num_leaves* sebanyak 31, dan *min_data_in_leaf* sebesar 20. Model tersebut menghasilkan nilai MAE sebesar 1.85, RMSE sebesar 2.76, dan MAPE sebesar 10.42%.

Nilai RMSE sebesar 2.76 menunjukkan adanya penurunan error dibandingkan dengan skenario A yang memiliki RMSE sebesar 3.03. Hal ini mengindikasikan bahwa peningkatan proporsi data latih memberikan dampak positif terhadap kemampuan model dalam memprediksi jumlah pengunjung. Nilai MAE dan MAPE yang juga lebih kecil menunjukkan bahwa selisih antara nilai aktual dan prediksi semakin berkurang, sehingga model menjadi lebih akurat.

Untuk memperkuat hasil evaluasi kuantitatif yang telah diperoleh, dilakukan analisis secara visual melalui grafik perbandingan antara data aktual dan hasil prediksi. Visualisasi ini bertujuan untuk melihat secara langsung kemampuan model dalam mengikuti pola pergerakan jumlah pengunjung serta mengidentifikasi adanya penyimpangan atau deviasi yang mungkin tidak terlihat hanya dari nilai metrik evaluasi.



Gambar 4.6 Grafik Perbandingan Data Aktual dan Prediksi Skenario B

Berdasarkan Gambar 4.6, data aktual jumlah pengunjung pada skenario B (80:20) menunjukkan pola fluktuatif yang cukup tinggi dengan beberapa lonjakan tajam hingga mendekati 100 pengunjung serta penurunan signifikan pada periode tertentu. Variasi ini mengindikasikan bahwa jumlah pengunjung harian memiliki dinamika yang tidak stabil, sehingga membutuhkan model yang mampu menangkap perubahan pola secara adaptif berdasarkan data historis yang tersedia.

Pada Gambar 4.6 juga terlihat bahwa garis prediksi mampu mengikuti pola umum dari data aktual dengan cukup baik. Model dapat merepresentasikan tren naik dan turun jumlah pengunjung, termasuk pada beberapa periode dengan lonjakan tinggi. Namun demikian, seperti pada skenario sebelumnya, garis prediksi tampak lebih halus dibandingkan data aktual sehingga pada beberapa titik perubahan yang sangat tajam masih terdapat sedikit selisih antara nilai prediksi dan nilai sebenarnya.

Meskipun masih terdapat deviasi kecil pada beberapa titik ekstrem, secara keseluruhan pola prediksi pada Gambar 4.6 terlihat lebih konsisten mengikuti pergerakan data aktual dibandingkan dengan skenario A. Pada skenario sebelumnya, garis prediksi cenderung lebih halus sehingga beberapa lonjakan tajam belum dapat diikuti dengan baik, sedangkan pada skenario B model mampu menangkap lebih banyak perubahan mendadak, termasuk lonjakan tinggi dan penurunan cepat yang terjadi pada beberapa periode. Selain itu, jarak antara garis prediksi dan data aktual pada nilai menengah hingga tinggi terlihat lebih rapat dibandingkan skenario 70:30.

Perbedaan ini menunjukkan bahwa penggunaan data latih yang lebih besar pada skenario 80:20 memberikan informasi historis yang lebih banyak bagi model untuk mempelajari pola fluktuasi data. Akibatnya, model menjadi lebih adaptif terhadap perubahan nilai yang ekstrem serta mampu merepresentasikan dinamika jumlah pengunjung dengan lebih baik. Hal ini juga didukung oleh nilai MAE, RMSE, dan MAPE yang lebih kecil dibandingkan skenario sebelumnya, sehingga dapat disimpulkan bahwa performa model pada skenario B lebih akurat dalam memprediksi jumlah pengunjung harian.

4.4.3 Skenario Pengujian C

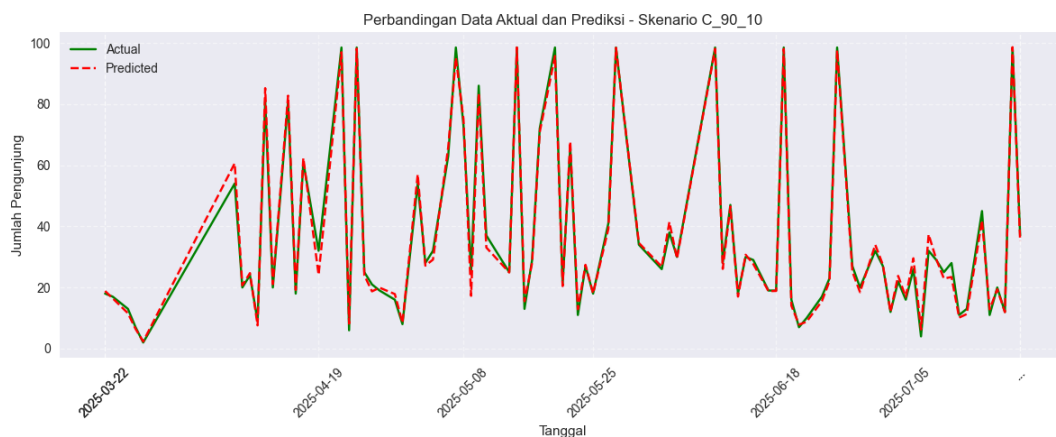
Pada skenario pengujian C, dataset dibagi menjadi 90% data latih dan 10% data uji berdasarkan urutan waktu. Proporsi ini memberikan porsi data latih paling besar dibandingkan skenario sebelumnya, sehingga model memiliki lebih banyak informasi historis untuk mempelajari pola fluktuasi jumlah pengunjung sebelum diuji pada periode akhir.

Berdasarkan hasil pengujian, diperoleh kombinasi hyperparameter terbaik dengan *learning_rate* sebesar 0,05, *n_estimators* sebanyak 300, *max_depth* sebesar 7, *num_leaves* sebanyak 15, dan *min_data_in_leaf* sebesar 20. Model tersebut menghasilkan nilai MAE sebesar 1,58, RMSE sebesar 2,27, dan MAPE sebesar 8,91%.

Nilai RMSE sebesar 2,27 merupakan yang paling kecil dibandingkan skenario A dan B. Hal ini menunjukkan bahwa model pada skenario C memiliki tingkat kesalahan prediksi paling rendah. Penurunan nilai MAE dan MAPE juga mengindikasikan bahwa selisih antara nilai aktual dan prediksi semakin kecil,

sehingga model mampu memberikan estimasi yang lebih mendekati kondisi sebenarnya.

Berdasarkan grafik perbandingan antara data aktual dan hasil prediksi pada Gambar 4.8, terlihat bahwa garis prediksi mengikuti pola aktual dengan lebih konsisten. Fluktuasi jumlah pengunjung dapat ditangkap dengan baik oleh model, termasuk pada beberapa perubahan tren yang terjadi. Tingkat deviasi yang lebih kecil dibandingkan dua skenario sebelumnya menunjukkan bahwa peningkatan proporsi data latih memberikan kontribusi positif terhadap performa model.



Gambar 4.7 Grafik Perbandingan Data Aktual dan Prediksi Skenario C

Berdasarkan Gambar 4.7, data aktual jumlah pengunjung pada skenario C (90:10) menunjukkan pola fluktuatif yang cukup tinggi dengan beberapa lonjakan tajam yang mendekati 100 pengunjung serta penurunan signifikan pada periode tertentu. Variasi ini menunjukkan bahwa jumlah pengunjung harian memiliki dinamika yang tidak stabil, sehingga diperlukan model yang mampu memahami pola perubahan secara lebih mendalam berdasarkan data historis yang tersedia.

Pada Gambar 4.7 juga terlihat bahwa garis prediksi (merah putus-putus) mampu mengikuti pola umum dari data aktual dengan sangat baik. Model dapat

merepresentasikan tren kenaikan dan penurunan jumlah pengunjung, termasuk pada beberapa periode dengan lonjakan yang cukup tinggi. Dibandingkan dengan skenario sebelumnya, garis prediksi pada skenario ini tampak lebih mendekati pola aktual, meskipun pada beberapa titik ekstrem masih terdapat sedikit perbedaan akibat perubahan yang sangat tajam.

Meskipun masih terdapat deviasi kecil pada beberapa titik, secara keseluruhan pola prediksi terlihat lebih konsisten mengikuti pergerakan data aktual. Dibandingkan dengan skenario A dan B, garis prediksi pada skenario 90:10 tampak lebih rapat terhadap garis aktual, terutama pada bagian dengan fluktuasi sedang hingga tinggi. Pada skenario A, model masih terlihat lebih halus sehingga beberapa lonjakan tajam belum sepenuhnya diikuti, sedangkan pada skenario B kemampuan model mulai membaik namun masih terdapat selisih pada perubahan ekstrem. Pada skenario C, model mampu mengikuti lebih banyak titik perubahan dengan deviasi yang lebih kecil, sehingga pola prediksi terlihat paling mendekati data aktual.

Kesesuaian ini menunjukkan bahwa peningkatan proporsi data latih pada skenario 90:10 memberikan kontribusi yang signifikan dalam membantu model mempelajari pola historis dengan lebih baik. Hal ini juga sejalan dengan nilai MAE, RMSE, dan MAPE yang paling rendah dibandingkan skenario sebelumnya, sehingga dapat disimpulkan bahwa model pada skenario C memiliki performa terbaik dalam memprediksi jumlah pengunjung harian.

4.5 Pembahasan

Bagian ini membahas hasil pemodelan yang telah dilakukan menggunakan metode LightGBM. Pembahasan difokuskan pada empat aspek utama, yaitu

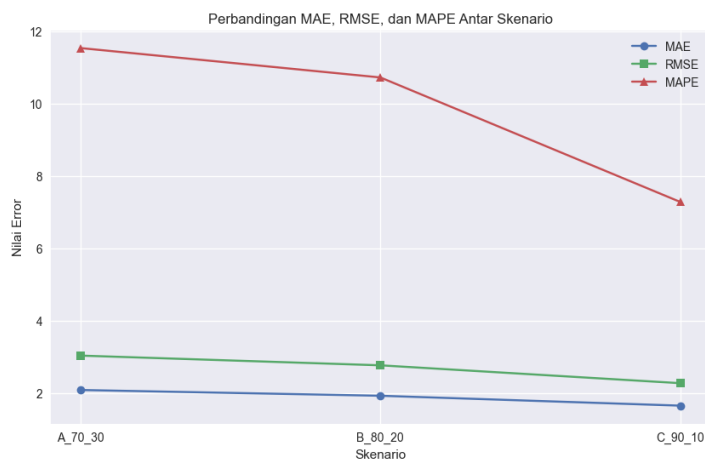
perbandingan skenario pembagian data, analisis hyperparameter, analisis metrik evaluasi, serta kesesuaian metode LightGBM dalam memprediksi jumlah pengunjung. Analisis ini bertujuan untuk menilai performa model secara menyeluruh dan melihat sejauh mana metode yang digunakan mampu menangkap pola data deret waktu.

4.5.1 Analisis Perbandingan Skenario Pembagian Data

Pada penelitian ini dilakukan tiga skenario pengujian berdasarkan perbedaan proporsi pembagian data latih (training) dan data uji (testing), yaitu 70:30 (Skenario A), 80:20 (Skenario B), dan 90:10 (Skenario C). Perbedaan proporsi ini bertujuan untuk menganalisis pengaruh jumlah data latih terhadap performa model LightGBM dalam memprediksi jumlah pengunjung harian.

Berdasarkan hasil evaluasi yang telah diperoleh, terjadi penurunan nilai error secara konsisten seiring dengan peningkatan proporsi data training. Pada Skenario A (70:30), model menghasilkan nilai MAE sebesar 2.08, RMSE sebesar 3.03, dan MAPE sebesar 11.54%. Ketika proporsi data latih ditingkatkan menjadi 80% pada Skenario B, nilai error menurun menjadi MAE 1.92, RMSE 2.76, dan MAPE 10.73%. Performa terbaik diperoleh pada Skenario C (90:10) dengan MAE sebesar 1.65, RMSE sebesar 2.27, dan MAPE sebesar 7.29%.

Penurunan nilai error tersebut menunjukkan bahwa semakin besar proporsi data latih yang digunakan, semakin baik kemampuan model dalam mempelajari pola fluktuasi jumlah pengunjung dari data historis. Dengan jumlah data pelatihan yang lebih banyak, model memperoleh informasi yang lebih lengkap untuk mengenali tren, perubahan nilai, serta variasi kunjungan harian secara lebih akurat.



Gambar 4.8 Grafik Perbandingan Evaluasi Model

Berdasarkan Gambar 4.9, terlihat bahwa ketiga metrik evaluasi menunjukkan tren penurunan dari Skenario A hingga Skenario C. Penurunan paling signifikan terlihat pada nilai MAPE, yang menunjukkan peningkatan akurasi model secara persentase. Secara keseluruhan, Skenario C menghasilkan nilai error terendah dibandingkan dua skenario lainnya, yang menunjukkan bahwa penggunaan data latih sebesar 90% memberikan performa prediksi yang paling baik.

Namun demikian, meskipun Skenario C menghasilkan nilai error terendah, proporsi data uji yang hanya sebesar 10% perlu diperhatikan. Jumlah data uji yang lebih sedikit dapat mempengaruhi representativitas evaluasi model. Oleh karena itu, pemilihan skenario terbaik tidak hanya mempertimbangkan nilai error terkecil, tetapi juga keseimbangan antara kemampuan pembelajaran model dan validitas pengujian. Secara keseluruhan, hasil pengujian menunjukkan bahwa peningkatan proporsi data latih berpengaruh positif terhadap performa model LightGBM dalam memprediksi jumlah pengunjung harian.

4.5.2 Analisis Hyperparameter

Pada penelitian ini dilakukan proses tuning hyperparameter untuk memperoleh konfigurasi model LightGBM yang menghasilkan performa terbaik pada masing-masing skenario pembagian data. Parameter yang dituning meliputi *learning_rate*, *n_estimators*, *max_depth*, *num_leaves*, dan *min_data_in_leaf*.

Perbandingan konfigurasi hyperparameter pada masing-masing skenario dapat dilihat pada Tabel 4.11.

Tabel 4.8 Perbandingan Hyperparameter

	learning_rate	n_estimators	max_depth	num_leaves	min_data_in_leaves
Skenario A	0.05	300.00	5.00	31.00	20.00
Skenario B	0.05	300.00	5.00	31.00	20.00
Skenario C	0.05	300.00	7.00	15.00	20.00

Berdasarkan Tabel 4.11, dapat diketahui bahwa nilai *learning_rate*, *n_estimators*, dan *min_data_in_leaf* cenderung konsisten pada ketiga skenario. Hal ini menunjukkan bahwa model memerlukan tingkat pembelajaran yang stabil dengan jumlah estimator yang cukup besar untuk memperoleh performa optimal.

Nilai *learning_rate* sebesar 0.05 menunjukkan bahwa model melakukan pembelajaran secara bertahap (gradual learning). Nilai ini relatif kecil sehingga pembaruan model pada setiap iterasi berlangsung secara lebih halus dan membantu mengurangi risiko *overfitting*.

Jumlah *n_estimators* sebanyak 300 menunjukkan bahwa model membangun 300 pohon keputusan untuk memperbaiki kesalahan prediksi secara

bertahap. Kombinasi antara *learning_rate* yang kecil dan jumlah estimator yang cukup besar memungkinkan model mencapai konvergensi yang lebih stabil.

Perbedaan yang paling signifikan terlihat pada parameter *max_depth* dan *num_leaves*. Pada Skenario A dan B, nilai *max_depth* sebesar 5 dengan *num_leaves* sebesar 31. Sementara itu, pada Skenario C nilai *max_depth* meningkat menjadi 7 namun *num_leaves* menurun menjadi 15.

Peningkatan *max_depth* memungkinkan model membangun struktur pohon yang lebih dalam sehingga mampu menangkap pola yang lebih kompleks dalam data kunjungan harian. Namun, pengurangan jumlah *num_leaves* membantu membatasi jumlah percabangan yang terbentuk, sehingga model tidak menjadi terlalu kompleks dan tetap memiliki kemampuan generalisasi yang baik.

Kombinasi tersebut menunjukkan bahwa performa terbaik tidak selalu diperoleh dari model dengan kompleksitas tertinggi, melainkan dari keseimbangan antara kedalaman pohon dan jumlah daun yang terbentuk. Hal ini tercermin pada Skenario C yang menghasilkan nilai error terendah dibandingkan dua skenario lainnya.

Parameter *min_data_in_leaf* sebesar 20 berperan dalam menjaga stabilitas model dengan memastikan bahwa setiap daun memiliki jumlah data minimum yang cukup, sehingga model tidak terlalu sensitif terhadap fluktuasi kecil atau noise dalam data.

Dengan demikian, dapat disimpulkan bahwa konfigurasi hyperparameter pada Skenario C merupakan kombinasi yang paling optimal dalam penelitian ini

karena mampu menghasilkan performa prediksi terbaik dengan tetap mempertahankan kemampuan generalisasi model.

4.5.3 Analisis Metrik Evaluasi

Evaluasi performa model pada penelitian ini menggunakan tiga metrik utama, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Ketiga metrik ini digunakan untuk memberikan gambaran tingkat kesalahan prediksi model dari berbagai sudut pandang, sehingga performa model dapat dianalisis secara lebih komprehensif.

MAE digunakan untuk mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi. Nilai MAE yang lebih kecil menunjukkan bahwa prediksi model semakin mendekati nilai sebenarnya. Pada penelitian ini, nilai MAE yang diperoleh berada pada kisaran yang relatif rendah, yang menunjukkan bahwa selisih antara nilai prediksi dan nilai aktual tidak terlalu besar.

RMSE memberikan penalti yang lebih besar terhadap kesalahan prediksi yang memiliki nilai tinggi karena proses kuadrat sebelum dirata-ratakan. Oleh karena itu, RMSE sensitif terhadap error ekstrem atau lonjakan nilai yang tidak dapat diprediksi dengan baik oleh model. Nilai RMSE yang tidak jauh berbeda dengan MAE pada penelitian ini menunjukkan bahwa model tidak menghasilkan banyak kesalahan besar, sehingga prediksi yang dihasilkan relatif stabil.

MAPE digunakan untuk mengukur tingkat kesalahan dalam bentuk persentase terhadap nilai aktual. Metrik ini memudahkan interpretasi karena menunjukkan seberapa besar kesalahan prediksi dibandingkan nilai sebenarnya. Nilai MAPE yang diperoleh pada penelitian ini berada di bawah 10% pada skenario

terbaik, yang menunjukkan bahwa model memiliki tingkat akurasi yang sangat baik dalam memprediksi jumlah pengunjung harian.

Berdasarkan ketiga metrik evaluasi tersebut, dapat diketahui bahwa model mampu menghasilkan prediksi yang cukup akurat dengan tingkat kesalahan yang relatif kecil. Kombinasi nilai MAE yang rendah, RMSE yang stabil, serta MAPE yang kecil menunjukkan bahwa model tidak hanya akurat secara rata-rata, tetapi juga mampu menjaga konsistensi prediksi terhadap fluktuasi data jumlah pengunjung harian.

4.5.4 Analisis Kesesuaian Metode *LightGBM*

Berdasarkan hasil evaluasi yang telah dilakukan, dapat disimpulkan bahwa metode *LightGBM* memiliki kemampuan yang baik dalam menangkap pola kompleks pada data deret waktu jumlah pengunjung. Hal ini terlihat dari kemampuannya dalam mengikuti fluktuasi data aktual, termasuk perubahan tren dan variasi nilai yang dinamis. Dengan memanfaatkan algoritma berbasis gradient boosting, *LightGBM* mampu membangun model yang adaptif terhadap pola historis tanpa mengalami overfitting yang signifikan.

Selain itu, penggunaan fitur-fitur tambahan seperti *lag* dan *rolling mean* turut berkontribusi dalam meningkatkan performa model. Fitur-fitur tersebut membantu model dalam memahami hubungan temporal antar data, sehingga prediksi yang dihasilkan menjadi lebih akurat dan stabil. Hal ini didukung oleh nilai MAE, RMSE, dan MAPE yang relatif kecil, yang menunjukkan bahwa model mampu meminimalkan kesalahan prediksi secara konsisten.

Dari sisi efisiensi, LightGBM juga memiliki keunggulan dalam proses pelatihan yang lebih cepat dan penggunaan memori yang lebih optimal dibandingkan metode boosting lainnya. Meskipun penelitian ini menggunakan dataset dengan skala terbatas, hasil yang diperoleh tetap menunjukkan performa yang optimal dan stabil pada berbagai skenario pengujian.

Lebih lanjut, konsistensi penurunan nilai MAE, RMSE, dan MAPE dari Skenario A hingga Skenario C menunjukkan bahwa LightGBM mampu memanfaatkan peningkatan jumlah data latih secara efektif. Hal ini menandakan bahwa model memiliki kemampuan generalisasi yang baik serta responsif terhadap penambahan informasi historis.

Dengan demikian, dapat disimpulkan bahwa metode LightGBM sangat sesuai digunakan dalam kasus prediksi jumlah pengunjung harian. Model ini tidak hanya mampu menghasilkan tingkat akurasi yang tinggi, tetapi juga memiliki stabilitas yang baik dalam menghadapi fluktuasi data. Oleh karena itu, LightGBM dapat direkomendasikan sebagai metode yang efektif untuk permasalahan prediksi berbasis data time series.

4.6 Intergrasi Islam

Penelitian prediksi jumlah pengunjung harian menggunakan metode LightGBM dapat diintegrasikan dengan nilai-nilai Islam yang menekankan pentingnya perencanaan, pengambilan keputusan yang bijak, serta profesionalitas dalam bekerja. Dalam ajaran Islam, manusia dianjurkan untuk mempersiapkan masa depan dengan mempertimbangkan kondisi yang mungkin terjadi melalui pemanfaatan informasi yang tersedia sebagai dasar dalam menyusun strategi yang

tepat. Hal ini menunjukkan bahwa upaya memahami pola masa lalu untuk memperkirakan kondisi di masa depan merupakan bagian dari ikhtiar manusia dalam menghadapi ketidakpastian.

Dalam konteks penelitian ini, prediksi jumlah pengunjung harian dilakukan menggunakan data historis untuk memperkirakan kondisi di masa mendatang. Hasil prediksi tersebut dapat membantu pengelola museum dalam menyusun strategi pengelolaan yang lebih terencana, seperti pengaturan fasilitas, sumber daya, serta pelayanan kepada pengunjung. Dengan demikian, proses prediksi tidak hanya berfungsi sebagai alat analisis, tetapi juga sebagai bentuk perencanaan yang sistematis.

Allah Swt. juga berfirman dalam Q.S. Al-Anfal ayat 60 :

وَأَعِدُّوا لَهُمْ مَا اسْتَطَعْتُمْ مِنْ قُوَّةٍ وَمِنْ رِبَاطِ الْخَيْلِ تُرْهِمُونَ بِهِ ۚ عَدُوَّ اللَّهِ وَعَدُوَّكُمْ وَآخَرِينَ مِنْ دُونِهِمْ لَا تَعْلَمُونَهُمُ اللَّهُ يَعْلَمُهُمْ ۚ وَمَا تُنْفِقُوا مِنْ شَيْءٍ فِي سَبِيلِ اللَّهِ يُوَفَّ إِلَيْكُمْ وَأَنْتُمْ لَا تُظْلَمُونَ ٦٠

“Dan persiapkanlah untuk menghadapi mereka kekuatan apa saja yang kamu sanggupi dan dari pasukan berkuda yang ditambat untuk berperang, yang dengan persiapan itu kamu menggentarkan musuh Allah, musuhmu dan orang-orang selain mereka yang kamu tidak mengetahuinya; Allah mengetahuinya. Apa saja yang kamu infakkan di jalan Allah niscaya akan dibalas dengan cukup kepadamu dan kamu tidak akan dizalimi.” (Q.S. Al-Anfal [8]: 60).

Menurut tafsir Ibnu Katsir, ayat ini menjelaskan bahwa umat Islam diperintahkan untuk mempersiapkan segala bentuk kekuatan yang mampu dilakukan guna menghadapi berbagai ancaman. Kekuatan tersebut tidak hanya berupa persenjataan, tetapi juga mencakup kemampuan, strategi, dan kesiapan yang dapat memberikan perlindungan serta menghadapi tantangan dengan baik.

Menurut tafsir M. Quraish Shihab dalam *Tafsir Al-Mishbah*, ayat ini menekankan pentingnya melakukan persiapan secara maksimal sebelum menghadapi suatu keadaan. Persiapan tidak hanya berkaitan dengan kekuatan fisik, tetapi juga meliputi strategi, kemampuan, dan perencanaan yang matang agar tujuan dapat tercapai secara optimal.

Sementara itu, dalam *Tafsir Ath-Thabari* dijelaskan bahwa perintah untuk mempersiapkan kekuatan menunjukkan pentingnya usaha dan kesiapan dalam menghadapi berbagai tantangan. Persiapan yang baik menjadi bentuk ikhtiar manusia agar mampu mengambil keputusan yang tepat serta menghindari kerugian atau kelemahan di masa mendatang.

Ayat dan tafsir tersebut memiliki keterkaitan dengan penelitian ini, yaitu pentingnya persiapan dan perencanaan sebelum mengambil keputusan. Dalam penelitian prediksi jumlah pengunjung menggunakan metode LightGBM, hasil prediksi dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan, seperti perencanaan fasilitas, pengaturan sumber daya, dan penyusunan strategi pelayanan. Dengan adanya prediksi, pihak terkait dapat mempersiapkan kebutuhan secara lebih efektif sehingga risiko kesalahan keputusan dapat diminimalkan.

Selanjutnya, Allah Swt. juga berfirman dalam Q.S. Al-Hasyr ayat 18:

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِإِعَادٍ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ ١٨

“Wahai orang-orang yang beriman! Bertakwalah kepada Allah dan hendaklah setiap diri memperhatikan apa yang telah diperbuatnya untuk hari esok, dan bertakwalah kepada Allah. Sesungguhnya Allah Maha Mengetahui apa yang kamu kerjakan.” (Q.S. Al-Hasyr [59]: 18).

Menurut tafsir M. Quraish Shihab dalam *Tafsir Al-Mishbah*, ayat ini menjelaskan bahwa setiap manusia diperintahkan untuk memperhatikan dan mengevaluasi apa yang telah dilakukan sebagai persiapan menghadapi masa depan. Perintah untuk memperhatikan “hari esok” menunjukkan pentingnya perencanaan, kehati-hatian, dan pertimbangan yang matang dalam setiap tindakan agar memperoleh hasil yang baik di kemudian hari.

Menurut tafsir Al-Qurthubi dalam *Tafsir Qurthubi*, ayat ini mengandung anjuran agar manusia senantiasa melakukan introspeksi terhadap amal perbuatannya serta mempersiapkan bekal terbaik untuk masa yang akan datang. Hal tersebut menunjukkan pentingnya sikap waspada, perhitungan, dan kesiapan sebelum mengambil suatu tindakan atau keputusan.

Sementara itu, dalam *Tafsir Ath-Thabari*, Ibnu Jarir Ath-Thabari menjelaskan bahwa ayat ini merupakan perintah bagi manusia untuk memperhatikan apa yang telah dipersiapkan untuk kehidupan di masa depan. Ayat ini menekankan pentingnya usaha dan pertimbangan yang baik dalam setiap tindakan agar manusia dapat menghadapi masa mendatang dengan lebih siap dan terarah.

Ayat dan tafsir tersebut memiliki keterkaitan dengan penelitian ini, yaitu pentingnya perencanaan dan pertimbangan sebelum mengambil keputusan. Dalam penelitian prediksi jumlah pengunjung menggunakan metode LightGBM, hasil prediksi dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan, seperti perencanaan fasilitas, pengaturan sumber daya, dan penyusunan strategi pelayanan museum. Dengan adanya prediksi berdasarkan data historis, pihak pengelola dapat

mempertimbangkan berbagai kemungkinan kondisi di masa mendatang sehingga keputusan yang diambil menjadi lebih efektif dan tepat sasaran.

Rasulullah saw. bersabda:

إِنَّ اللَّهَ يُحِبُّ إِذَا عَمِلَ أَحَدُكُمْ عَمَلًا أَنْ يُتَّقِنَهُ

“Sesungguhnya Allah mencintai apabila salah seorang di antara kalian melakukan suatu pekerjaan, ia menyempurnakannya.” (HR. Thabrani).

Menurut penjelasan para ulama, hadis ini mengandung prinsip *itqan*, yaitu melakukan pekerjaan secara sungguh-sungguh, teliti, dan profesional agar memperoleh hasil yang terbaik. Setiap pekerjaan hendaknya dilakukan dengan penuh tanggung jawab, ketepatan, serta memanfaatkan kemampuan secara optimal sehingga pekerjaan tersebut dapat memberikan manfaat yang maksimal.

Menurut Abu Al Qasim Ath-Thabarani dalam kitab *Al Mu'jam Ash-Shaghir*, hadis ini menunjukkan bahwa Allah Swt. menyukai pekerjaan yang dilakukan dengan baik dan sempurna. Sikap profesional, ketelitian, dan kesungguhan dalam menyelesaikan pekerjaan merupakan bentuk usaha yang dianjurkan dalam Islam agar hasil yang diperoleh menjadi lebih optimal.

Sementara itu, para ulama juga menjelaskan bahwa konsep *itqan* tidak hanya berkaitan dengan ibadah, tetapi juga mencakup pekerjaan dan aktivitas sehari-hari, termasuk dalam bidang ilmu pengetahuan dan teknologi. Oleh karena itu, setiap proses yang dilakukan perlu memperhatikan kualitas, evaluasi, dan perbaikan agar hasil yang diperoleh lebih akurat dan dapat dipertanggungjawabkan.

Hadis dan penjelasan tersebut memiliki keterkaitan dengan penelitian ini, yaitu pentingnya bekerja secara profesional dan optimal dalam melakukan prediksi

jumlah pengunjung museum. Dalam penelitian ini, proses prediksi dilakukan menggunakan metode LightGBM yang disertai dengan *tuning hyperparameter* serta evaluasi menggunakan MAE, RMSE, dan MAPE. Langkah-langkah tersebut menunjukkan upaya sistematis untuk menghasilkan model prediksi yang lebih akurat. Dengan demikian, penggunaan metode prediksi dalam penelitian ini mencerminkan penerapan prinsip profesionalitas dan ketelitian dalam mengolah data guna membantu pengambilan keputusan yang lebih efektif di masa mendatang.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode LightGBM mampu digunakan untuk memprediksi jumlah pengunjung harian dengan tingkat akurasi yang cukup baik. Proses preprocessing dan feature engineering, seperti ekstraksi fitur waktu, fitur shock, lag, dan rolling mean, berperan penting dalam membantu model memahami pola historis data.

Hasil pengujian dengan tiga skenario pembagian data menunjukkan bahwa proporsi data latih berpengaruh terhadap performa model. Skenario 90:10 menghasilkan performa terbaik dengan nilai RMSE sebesar 2,27, MAE sebesar 1,58, dan MAPE sebesar 8,91%. Nilai tersebut menunjukkan bahwa rata-rata kesalahan prediksi model berada pada kisaran ± 2 pengunjung dari nilai aktual.

Secara keseluruhan, peningkatan jumlah data latih terbukti mampu meningkatkan kemampuan model dalam menangkap pola fluktuasi jumlah pengunjung. Dengan demikian, penggunaan metode LightGBM dengan kombinasi feature engineering yang tepat dapat menjadi pendekatan yang efektif dalam peramalan jumlah pengunjung harian.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian selanjutnya.

- a. Penelitian selanjutnya dapat menambahkan variabel eksternal seperti hari libur nasional, cuaca, atau event tertentu yang berpotensi mempengaruhi jumlah pengunjung sehingga model dapat menghasilkan prediksi yang lebih akurat.
- b. Pengujian dapat dilakukan dengan membandingkan metode lain, seperti XGBoost, Random Forest, atau model berbasis deep learning seperti LSTM, untuk mengetahui metode yang paling optimal dalam kasus prediksi jumlah pengunjung.
- c. Pengembangan sistem dapat diarahkan pada implementasi model dalam bentuk sistem monitoring atau dashboard prediksi secara real-time agar hasil penelitian dapat dimanfaatkan secara langsung oleh pengelola.

DAFTAR PUSTAKA

- Abu Ja'far Muhammad bin Jarir ath-Thabari. (n.d.). *Tafsir Ath-Thabari*. Pustaka Azzam.
- Agustin, Y. H., Satria, E., & Siti Nursifa, F. (2025). Prediksi Jumlah Pengunjung Pariwisata di Kabupaten Garut Menggunakan Algoritma Regresi Linear. *Jurnal Algoritma*, 22(1), 569–581. <https://doi.org/10.33364/algoritma/v.22-1.1807>
- Ardesfira, G., Zedha, H. F., Fazana, I., Rahmadhiyanti, J., Rahima, S., Anwar, S., Statistika, J., Kuala, U. S., Aceh, B., & Tukar, N. (2022). Jambura Journal Of Probability And Statistics Volume 3 Nomor 2, November 2022. *Jambura Journal of Probability and Statistics*, 3(November 2022).
- Arwansyah, A., Suryani, S., SY, H., Usman, U., Ahyuna, A., & Alam, S. (2022). Time Series Forecasting Menggunakan Deep Gated Recurrent Units. *Digital Transformation Technology*, 4(1), 410–416. <https://doi.org/10.47709/digitech.v4i1.4141>
- Aulady, M. A. (2024). *Improve Metode Lightgbm untuk Prediksi Harga Mobil Bekas Menggunakan Hyper-Parameter Tuning*. 5(September), 456–467. <https://doi.org/10.33650/trilogi.v5i3.9000>
- Ayu, I. G., Astarani, R., & Surya, I. G. (2025). Analisis Perbandingan XGBoost dan LightGBM dalam Prediksi Penjualan Ritel Walmart Store Sales. 3(MI), 717–728.
- Bombrun, M., Dash, J. P., Pont, D., Watt, M. S., Pearse, G. D., & Dungey, H. S. (2020). Forest-Scale Phenotyping: Productivity Characterisation Through Machine Learning. 11(March), 1–14. <https://doi.org/10.3389/fpls.2020.00099>
- Budiawan, I., Yasin, S., Harafani, H., Kiswanto, A. D., Rusli, A. R., & Marthanti, A. S. (2024). Analisis Data Kunjungan Wisatawan Mancanegara ke Indonesia pada Era Pasca Pandemi melalui Metode Visualisasi dan Peramalan. *Jurnal Pendidikan* ..., 8, 4787–4800. <https://www.jptam.org/index.php/jptam/article/view/13125%0Ahttps://www.jptam.org/index.php/jptam/article/download/13125/10079>
- Cherrly, A., & Somya, R. (2023). Prediksi Penjualan Tiket Wisata Taman Bermain Menggunakan Metode ARIMA. *Techno.Com*, 22(2), 312–322. <https://doi.org/10.33633/tc.v22i2.7950>
- Dahlan, M. (2024). Fungsi dan Peran Museum Sebagai Sumber Belajar dan Media Pelestarian Sejarah. *Jurnal Kajian Sosial Dan Budaya*, 8(April), 25–29. <https://www.ejournal.tebarscience.com/index.php/JKSB>
- Departemen Agama RI. (n.d.). *Al Qur'an dan Tafsirnya Jilid 4 Juz 10-11-12* (10th ed.). Widya Cahaya.

- Fahmi, A., Hafidz, A., Puspaningrum, E. Y., & Nurlaili, A. L. (2025). *Ruang Warna Hsv Untuk Klasifikasi Rempah Rimpang*. 9(3), 5492–5498.
- Ibnu Katsir. (n.d.). *Tafsir Ibnu Katsir Jilid 6*. Pustaka Imam Syafi’.
- Iskandar, Y. (2023). Analisis Peramalan Jumlah Wisatawan Geopark Ciletuh Menggunakan Exponential Smoothing. *Jurnal Informatika Teknologi Dan Sains*, 5(1), 165–169. <https://doi.org/10.51401/jinteks.v5i1.2222>
- Julianto, I. R., Indwiarti, & Rohmawati, A. A. (2021). Prediksi Jumlah Kunjungan Wisatawan Di Jawa Barat Dengan Model Arimax Dan Sarimax Menggunakan Data Google Trends. *E-Proceeding of Engineering*, 8(435), 4229–4241.
- Katsir, I. (n.d.). *Tafsir Ibnu Katsir Jilid 4*. Pustaka Imamm Syafi’i.
- Kurniadi, F. I., & Larasati, P. D. (2022). Light Gradient Boosting Machine untuk Deteksi Penyakit Stroke. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 6(1), 67–72. <https://doi.org/10.47970/siskom-kb.v6i1.328>
- Liang, S., Deng, T., Huang, A., Liu, N., & Jiang, X. (2023). Energy consumption prediction using the GRU-MMattention-LightGBM model with features of Prophet decomposition. *PLoS ONE*, 18(1 January). <https://doi.org/10.1371/journal.pone.0277085>
- Manullang, J., Santoso, A. J., & Emanuel, A. W. R. (2020). Prediksi Kunjungan Wisatawan Taman Nasional Gunung Merbabu dengan Time Series Forecasting dan LSTM. *Jurnal Buana Informatika*, 11(2), 132–140. <https://doi.org/10.24002/jbi.v11i2.3825>
- Meimela, A. (2021). Prediksi Jumlah Kunjungan Wisatawan Mancanegara ke Indonesia. *Media Wisata*, 19(1), 34–41. <https://doi.org/10.36276/mws.v19i1.64>
- Mohamad, S., Hasan, R., & Wantu, A. (2024). Optimalisasi Peran Museum sebagai Sumber Pelestarian Budaya dalam Pembelajaran Sejarah Lokal di Sekolah. *SOSIOLOGI: Jurnal Penelitian Dan Pengabdian Kepada Masyarakat*, 1(3), 197–202.
- Najib, F. A., & Nafi’iyah, N. (2020). Algoritma SVM untuk Memprediksi Pengunjung Wisata Musium di Jakarta Article History ABSTRAK. *Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika (KERNEL)*, 1(1), 24–32.
- Nullhakim, P. A., Chairunisah, C., & Arnita, A. (2024). Prediksi Jumlah Wisatawan Mancanegara Ke Sumatera Utara Berdasarkan Pintu Masuk Utama Menggunakan Algoritma Backpropagation Neural Network. *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, 7(2), 78–89. <https://doi.org/10.47324/ilkominfo.v7i2.264>
- Nur Fauzi, N. P., Khomsah, S., & Putra Wicaksono, A. D. (2025). Penerapan

Feature Engineering dan Hyperparameter Tuning untuk Meningkatkan Akurasi Model Random Forest pada Klasifikasi Risiko Kredit. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 12(2), 251–262. <https://doi.org/10.25126/jtiik.2025128472>

Ramadhani, A., Wahyuningsih, S., & Siringoringo, M. (2022). Peramalan Jumlah Kunjungan Wisatawan Mancanegara Ke Indonesia Menggunakan Autoregressive Integrated Moving Average (ARIMA). *Eksponensial*, 13(2), 103. <https://doi.org/10.30872/eksponensial.v13i2.1049>

Saputra, M. A., & Sugihartono, T. (2025). *Evaluasi Kinerja Model LSTM Untuk Prediksi Risiko Penyakit Jantung Menggunakan Dataset Program Studi Teknik Informasi , Fakultas Teknologi Informasi , ISB Atma Luhur , Indonesia Performance Evaluation of the LSTM Model for Heart Disease Risk Prediction Us.* 5(7), 1823–1833.

Septiana Rizky, P., Haiban Hirzi, R., & Hidayaturrohman, U. (2022). Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 15(2), 228–236. <https://doi.org/10.36456/jstat.vol15.no2.a5548>

Shihab, M. Q. (n.d.). *Tafsir Al Mishbah Vol.6*. Penerbit Lentera Hati.

Wang, Y., & Wang, T. (2020). *applied sciences Application of Improved LightGBM Model in Blood Glucose Prediction.*

Wibowo, D. B., & Tuti Purwaningsih. (2023). Analisis Peramalan Jumlah Pengunjung Daya Tarik Wisata Kabupaten Bantul dengan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA). *Emerging Statistics and Data Science Journal*, 1(1), 117–127. <https://doi.org/10.20885/esds.vol1.iss.1.art13>

Yuliansyah, H. (2025). *Peramalan Jumlah Pengunjung Wisata Edukasi Museum Menggunakan Kombinasi Moving Average Dan Model Prophet Program Studi Informatika , Universitas Ahmad Dahlan , Indonesia Forecasting the Number of Visitors to Educational Tourism Museums Using a Combination.* 5(7), 2033–2054.

Zihad, D. A. N., & Windari, S. (2025). *Koleksi sebagai Media Edukasi: Strategi Tata Kelola Museum yang Berorientasi Pembelajaran.* 8(8).