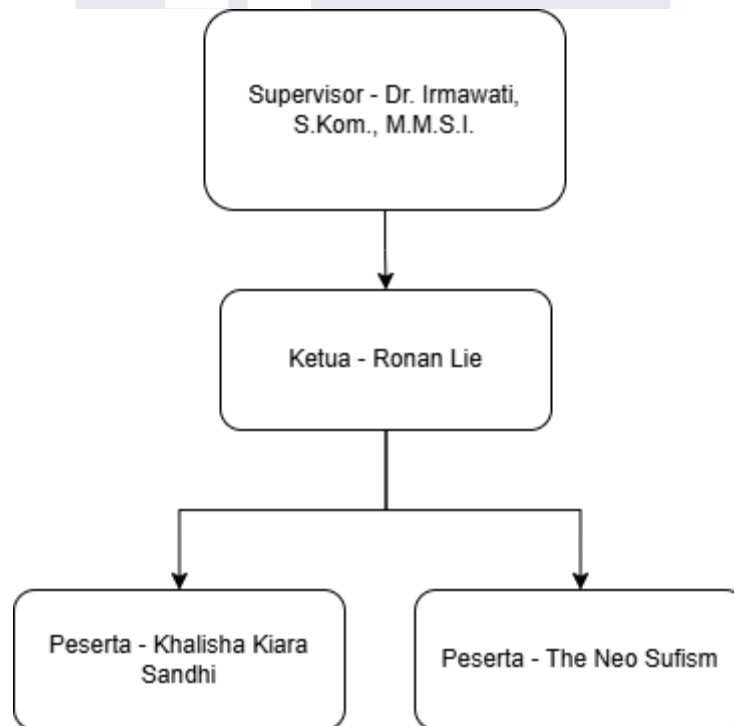


BAB III

PELAKSANAAN PRO-STEP : ROAD TO CHAMPION

3.1 Kedudukan dan Koordinasi



Gambar 3.1 Bagan Alur Koordinasi

Pada gambar 3.1 peserta bekerja dalam tim dengan Ronan Lie, The Neo Sufism, dan Khalisha Kiara Sandhi untuk menjalankan PRO-STEP: Road to Champion Program jalur kompetisi. Dr. Irmawati, S.Kom., M.M.S.I., bertindak sebagai Dosen Pembimbing Internal untuk tim ini dan memberikan bimbingan akademik tentang metodologi penelitian, penyusunan laporan, dan prosedur pelaksanaan proyek. Setiap koordinasi dengan pembimbing dilakukan secara

berkala untuk memastikan bahwa langkah-langkah tepat dan sudah teratur

Struktur kerja tim menetapkan tugas teknis untuk setiap anggota tim. Ronan Lie adalah penulis yang bertanggung jawab untuk memastikan bahwa seluruh tahapan yang berkaitan dengan algoritma LightGBM didokumentasikan dengan baik dan sesuai dengan persyaratan penelitian. Pengembangan algoritma LightGBM mencakup proses praproses data, pembangunan model, evaluasi, dan pengembangan skema optimasi model.

Selanjutnya, Neo Sufism bertanggung jawab atas pengembangan model Random Forest, serta alur kerja dan parameter yang digunakan. Peserta bertanggung jawab untuk menyesuaikan struktur pipeline, metrik evaluasi, dan format modeling agar relevan dengan penulisan ilmiah. Sementara itu, Khalisha Kiara Sandhi adalah pengembang model XGBoost dan bertanggung jawab atas proses pemodelan, validasi, dan pembuatan visualisasi yang diperlukan untuk penjelasan teknis.

Seluruh anggota tim bekerja sama dalam alur koordinasi untuk memastikan konsistensi antar model, kesesuaian metodologi, dan integrasi hasil modeling ke dalam kerangka penelitian. Diskusi internal dilakukan untuk menyelaraskan keputusan teknis, membagi jadwal pekerjaan, dan memastikan bahwa setiap anggota dapat menyelesaikan tugas sesuai jadwal. Sebelum melanjutkan ke tahap berikutnya, setiap perkembangan pekerjaan dikonsultasikan dengan Dosen Pembimbing Internal untuk mendapatkan arahan tambahan, klarifikasi metodologis, dan validasi akademik.

Seluruh proses pengerjaan proyek kompetisi dapat dilakukan dengan sistematis melalui pembagian tugas yang jelas dan koordinasi yang terstruktur ini, mulai dari pengolahan data, pengembangan model masing-masing algoritma,

hingga penyusunan laporan dan persiapan presentasi.



Gambar 3.2 Jobdesc peserta

Pada gambar 3.2 struktur kedudukan berikut menunjukkan hubungan kerja antara penulis, anggota tim, dan dosen pembimbing dalam pelaksanaan PRO-STEP: Road to Champion Program jalur kompetisi. Struktur ini dirancang untuk menunjukkan alur tanggung jawab, peran masing-masing, dan hubungan kerja selama proses penyusunan penelitian dan persiapan lomba.

Supervisor Dr. Irmawati, S.Kom., M.M.S.I. Supervisor atau Dosen Pembimbing Lomba berada di tingkat tertinggi struktur organisasi dan bertanggung jawab untuk memberikan arahan akademik dan memastikan bahwa seluruh proses penelitian dan kegiatan kompetisi dilakukan secara ilmiah. Salah satu tanggung jawab utama supervisor adalah memberikan bimbingan metodologis, menilai kelayakan metode penelitian, memberikan saran untuk perbaikan laporan dan presentasi, dan memantau perkembangan tim secara teratur. Supervisor juga bertanggung jawab untuk menjamin kualitas, memastikan bahwa setiap keputusan penting yang dibuat tim dibahas terlebih dahulu.

Ketua tim Ronan Lie bertanggung jawab untuk mengembangkan algoritma LightGBM secara teknis. Ini termasuk tahapan praproses data, pembangunan model, evaluasi, dan optimasi lanjutan menggunakan pengaturan hyperparameter. Ketua tim juga memastikan bahwa seluruh hasil kerja teknis terintegrasi dengan baik dalam laporan dan siap untuk dipresentasikan.

Peserta Khalisha Kiara Sandhi, anggota tim yang bertanggung jawab untuk mengembangkan model XGBoost. Pemodelan, analisis parameter dasar, dan pembuatan visualisasi pendukung adalah semua tugas Khalisha. Semua tugas ini diperlukan untuk menyusun laporan dan presentasi. Khalisha membantu dalam penulisan naskah ilmiah dan pembuatan slide presentasi untuk tahap akhir, selain tugas teknis.

Peserta The Neo Sufism adalah anggota tim yang bertanggung jawab atas pembangunan dan pengujian model Random Forest. Neo melakukan analisis fitur, praproses data lanjutan, dan dokumentasi parameter yang digunakan pada model tersebut. Neo juga terlibat dalam integrasi hasil modeling ke dalam laporan dan bekerja sama dengan anggota lain untuk memastikan metodologi dan penulisan naskah ilmiah konsisten

Pencatatan Rangkuman Mingguan Proses PRO-STEP: Road To Champion Program

Tabel 3.1 Detail Pekerjaan yang Dilakukan PRO-STEP : Road to Champion Program

No.	Waktu Pengerjaan	Proyek	Keterangan
1	15 September – 20 September	Pendaftaran	Melakukan pendaftaran lomba Carbon 7.0
2	20 September – 03 Oktober	Membuat Abstrak	Proses membuat abstrak menggunakan data GAIKINDO
3	03 Oktober	Registrasi & Pengumpulan abstrak	Mengumpulkan abstrak
4	04 Oktober	Pengumuman lulus abstrak	Lulus ke babak full paper
5	05 Oktober – 21 Oktober	Memilih model	Melakukan riset terdahulu untuk mencari model algoritma yang cocok
6	05 Oktober – 21 Oktober	Melakukan forecasting	Menggunakan LightGBM, XGBoost, dan RandomForest
7	05 Oktober – 21 Oktober	Mengerjakan artikel ilmiah	Mengerjakan full paper LKTIN
8	21 Oktober	Pengumpulan full paper	Melakukan pengumpulan full paper dan registrasi ulang
9	25 Oktober	Pengumuman 10 besar Finalis	Lulus ke babak final
10	06 November	Technical Meeting	Mengikuti acara technical meeting
11	07 November	Presentasi	Melakukan presentasi
12	09 November	Pengumuman Pemenang	Mendapat penghargaan “ <i>Best Presentation</i> ”
13	09 November – 15 Desember	<i>Post-competition</i>	Melakukan hypertuning untuk lightgbm

3.1 Uraian Pelaksanaan Kerja

3.1.1 Proses Pelaksanaan

Semua kegiatan selama mengikuti PRO-STEP: Road to Champion Program di jalur kompetisi atau lomba dibahas carbon 7.0 di bagian sini. Pengerjaan sesuai dengan buku panduan CARBON 07 kompetisi artikel ilmiah. Penggalan ide, pengolahan data, pembuatan model pembelajaran mesin, penulisan karya ilmiah, , persiapan presentasi dan pembuatan poster adalah semua kegiatan dalam rangkaian kegiatan *road to champion*. Proses perancangan dari awal hingga akhir adalah apa yang dilakukan selama program berjalan.

3.1.1.1 Pendaftaran dan Pengumpulan Abstrak

Pendaftaran dan pengumpulan abstrak adalah langkah pertama yang dilakukan. Pada tahap ini, peserta harus melakukan registrasi melalui sistem atau formulir yang disediakan oleh panitia CARBON 07 untuk memverifikasi bahwa mereka adalah peserta dan memiliki kelengkapan administrasi yang diperlukan. Setelah proses pendaftaran selesai, peserta kemudian mengirimkan abstrak penelitian yang menjelaskan latar belakang masalah, tujuan penelitian, metode yang digunakan, dan kontribusi ilmiah.

Abstrak berfungsi sebagai representasi kualitas dan garis besar penelitian yang akan dikembangkan pada tahapan selanjutnya, sehingga pengumpulan abstrak merupakan langkah awal yang sangat penting. Komite dan dewan juri akan menilai gagasan baru, relevansinya dengan tema lomba, dan potensi inovasi penelitian. Ini sekaligus menjadi titik awal seleksi awal untuk menentukan apakah penelitian layak dilanjutkan. Setelah itu, peserta akan mengembangkan metodologi, analisis data, dan hasil penelitian secara menyeluruh di babak penyusunan kertas lengkap.

Oleh karena itu, tahap pendaftaran dan pengumpulan abstrak tidak hanya

bersifat administratif, tetapi juga berfungsi sebagai penentu kualitas awal karya ilmiah yang akan dilombakan, dan memberikan kesempatan bagi peserta untuk menunjukkan relevansi dan inovatif penelitiannya.

3.1.1.2 Penyusunan Abstrak

Abstrak disusun berdasarkan analisis awal data penjualan mobil GAIKINDO, serta data sekunder seperti Google Trends, inflasi, nilai tukar, dan faktor eksternal lainnya yang dapat memengaruhi permintaan mobil di Indonesia. Pada tahap ini, proses peninjauan pola umum data dilakukan, menemukan tren historis, musiman (seasonality), dan hubungan potensial antara variabel prediktor dan penjualan mobil. Proses eksplorasi juga melibatkan analisis awal untuk memahami karakteristik dataset, menemukan anomali, dan menilai kelayakan variabel untuk digunakan dalam model prediksi.

Proses menganalisis data ini sangat penting untuk menentukan tujuan dan metode penelitian. Hasil analisis awal digunakan untuk menulis abstrak yang membahas latar belakang masalah, tujuan penelitian, metode yang akan digunakan, dan bagaimana penelitian ini dapat membantu industri otomotif. Oleh karena itu, abstrak berfungsi sebagai pengantar yang menjelaskan mengapa penelitian ini penting dan memberikan juri cara untuk menilai relevansinya sebelum melanjutkan ke tahap berikutnya, yaitu menulis makalah lengkap.

3.1.1.3 Registrasi dan Pengumpulan Abstrak

Bagian ini dari proses pendaftaran dan pengiriman abstrak kompetisi CARBON 07 merupakan bagian administratif. Peserta harus memastikan bahwa mereka telah memenuhi semua persyaratan, seperti menulis abstrak dalam format yang benar, memastikan dokumen lengkap, dan mengunggahnya tepat waktu sesuai batas waktu yang ditetapkan oleh panitia. Proses verifikasi ini penting untuk memastikan bahwa proposal penelitian memenuhi standar administratif kompetisi sebelum masuk ke tahap penilaian akademik.

Setelah tahap administratif ini, yang menentukan siapa yang dapat melanjutkan

ke tahap persiapan makalah lengkap, hanya abstrak yang lolos seleksi administratif dan penilaian awal juri yang dapat diubah menjadi makalah ilmiah lengkap. Oleh karena itu, fase penyaringan pertama bertujuan untuk memastikan bahwa peserta siap dan layak sebelum memulai proses penelitian yang lebih mendalam.

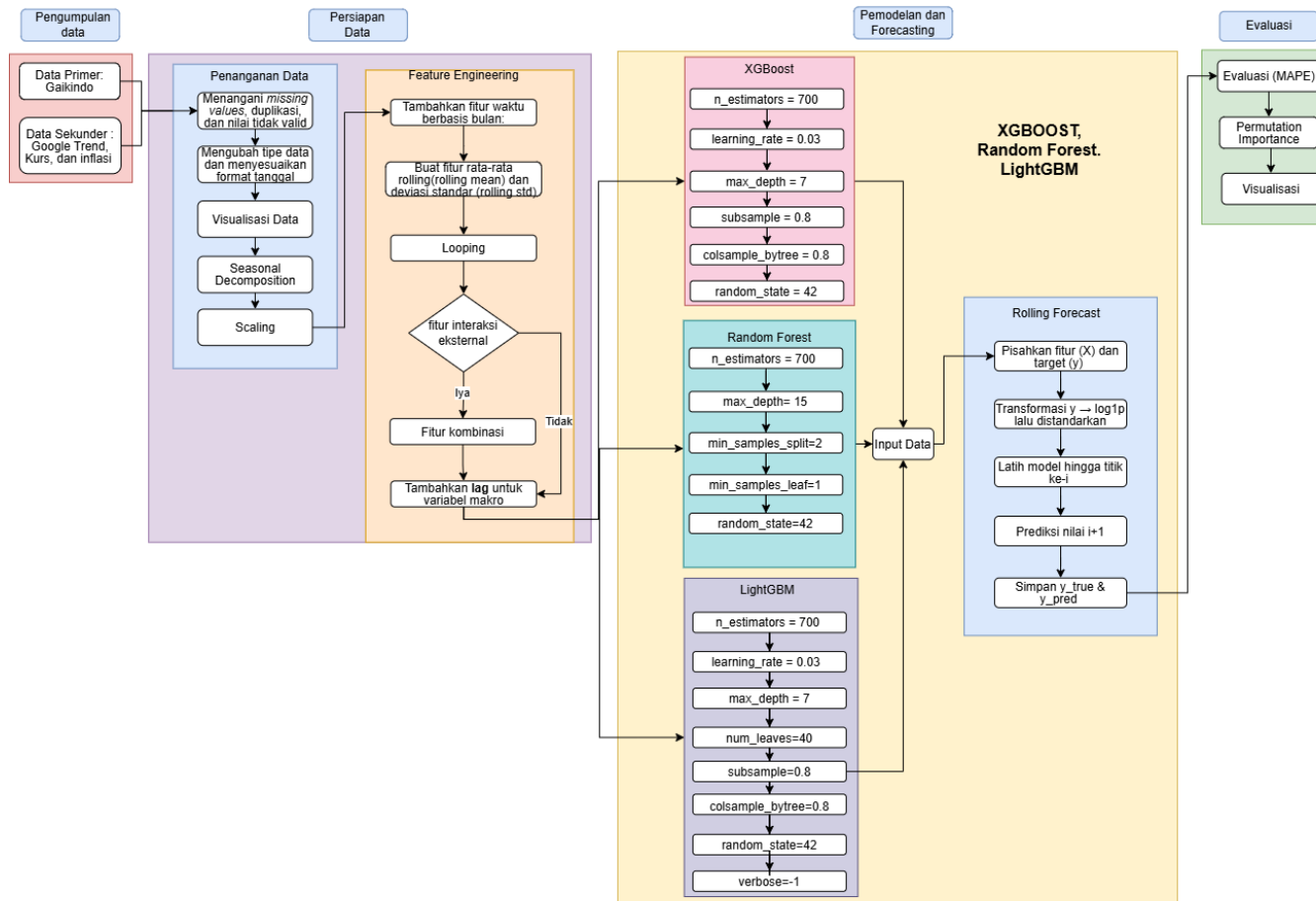
3.1.1.4 Pengumuman Kelulusan Abstrak

Komite Kompetisi CARBON 07 telah mengumumkan hasil seleksi abstrak kepada para peserta pada tahap ini. Peserta yang lolos diizinkan untuk melanjutkan ke tahap berikutnya, di mana mereka harus menulis makalah lengkap berdasarkan ide penelitian yang telah mereka tulis dalam abstrak. Proses ini menandai transisi ke fase penelitian yang lebih komprehensif. Pada tahap ini, peserta akan mulai menganalisis data secara sistematis dengan menggunakan metode penelitian yang mereka pilih dan membangun argumen ilmiah yang lebih terorganisir.

Selain meningkatkan dan memperkuat kontribusi teoretis, tahap ini memberi peserta kesempatan untuk memeriksa kredibilitas model atau metodologi penelitian yang diusulkan. Menang dalam kompetisi ilmiah nasional ini menunjukkan bahwa penelitian ini berkualitas tinggi dan memiliki potensi untuk menghasilkan gagasan baru.

3.1.1.5 Memilih Model Algoritma

Pada tahap ini, Anda perlu menulis artikel ilmiah lengkap berdasarkan semua hasil peramalan yang diperoleh dari penggunaan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Saat menulis artikel ilmiah lengkap, setiap langkah penelitian dijelaskan secara berurutan, mulai dari memahami bisnis dan data, mempersiapkan data, memodelkan data, hingga mengevaluasi hasilnya. Hal ini dilakukan untuk memastikan alur penelitian jelas dan membantu pembaca menilai kualitas proses analisis serta validitas hasil penelitian. Menulis artikel ilmiah lengkap merupakan langkah penting dalam menyusun cerita ilmiah yang utuh dari awal hingga akhir dan menjelaskan hasil penelitian secara menyeluruh.



Gambar 3.3 Metode Penelitian

Pada gambar 3.3 menunjukkan bagaimana proses penelitian dimulai dengan pengumpulan data dari dua sumber utama. Data primer dikumpulkan dari GAIKINDO tentang penjualan mobil nasional, dan data sekunder mencakup inflasi, kurs rupiah, dan Google Trends. Variabel eksternal digunakan untuk meningkatkan model. Karena kualitas data menentukan akurasi hasil prediksi, tahap awal ini membentuk dasar analisis. Persiapan data terjadi setelah semua data dikumpulkan. Proses ini mencakup sejumlah tindakan penting seperti menangani nilai yang hilang, menghilangkan duplikat, memperbaiki data yang tidak valid, menyesuaikan format tanggal, dan mengubah jenis data untuk memastikan bahwa data siap diproses oleh model.

Kemudian, metode dekomposisi musiman digunakan untuk melakukan visualisasi eksplorasi dan mencari pola musiman. Hal ini dilakukan untuk menemukan variabel acak, pola yang terjadi pada waktu-waktu tertentu dalam setahun, dan tren yang berlangsung dalam jangka waktu yang lama. Setelah menemukan pola data dasar, dilakukan standardisasi atau penskalaan untuk membuat rentang nilai antara variabel menjadi sama. Ini sangat penting untuk model berbasis machine learning untuk mencegah perbedaan antara fitur kecil dan besar.

Tahap berikutnya adalah proses feature engineering, atau proses pembuatan fitur baru, untuk meningkatkan kualitas prediksi. Pada tahap ini, variabel berbasis waktu seperti lag variabel makroekonomi, simpangan baku bergerak, bulan, dan rata-rata bergerak disertakan. Selain itu, fitur interaksi eksternal dibuat jika ada hubungan yang signifikan antara variabel prediktor dan variabel target. Teknik fitur sangat penting untuk meningkatkan kinerja model karena memungkinkan algoritma menangkap pola yang lebih kompleks dan kontekstual.

Tahap pemodelan menggunakan tiga algoritma pembelajaran mesin utama: XGBoost, Random Forest, dan LightGBM. Kami menyesuaikan hiperparameter masing-masing algoritma agar bekerja dengan optimal. Untuk memprediksi titik $i+1$, data dilatih hingga titik waktu i menggunakan metode peramalan bergulir. Untuk mengatasi distribusi yang tidak seimbang, fitur dan target dipisahkan, fungsi $\log1p$

digunakan untuk mengubah data target, dan kemudian data distandarisasi sebelum pelatihan. Setelah membandingkan nilai prediksi dengan nilai aktual, hasilnya disimpan untuk evaluasi kinerja.

Langkah terakhir adalah mengevaluasi model, yang menggunakan ukuran Mean Absolute Percentage Error (MAPE) untuk menunjukkan seberapa salah prediksi tersebut. Analisis pentingnya permutasi juga dilakukan untuk mengetahui fitur mana yang paling berpengaruh terhadap prediksi model. Selain itu, nilai prediksi dan nilai aktual ditampilkan dengan cara yang memudahkan untuk melihat perbedaan hasil model. Penilaian ini membantu orang memahami seberapa andal model tersebut dan mengapa model ini adalah yang terbaik untuk digunakan.

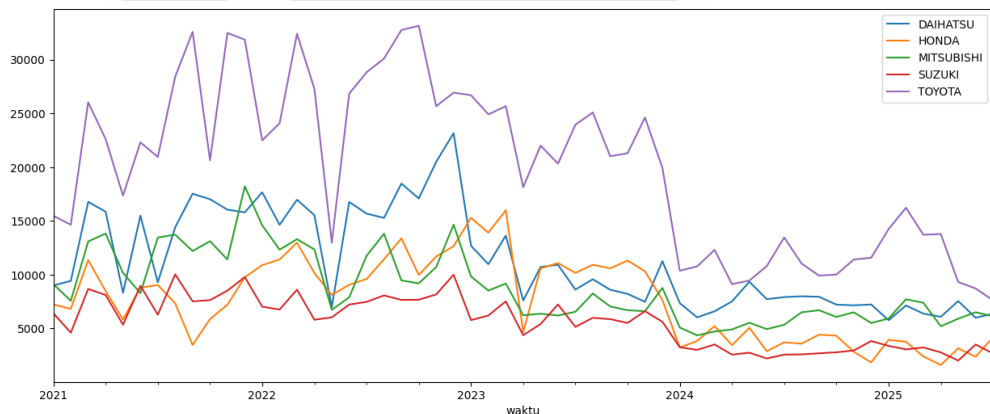
3.1.1.6 Proses Forecasting

Tahap Pemahaman Bisnis mengutamakan pemahaman konteks bisnis dan temuan penelitian. Produsen dan distributor dalam industri otomotif sangat membutuhkan prediksi penjualan mobil yang akurat karena ini berdampak langsung pada keputusan yang dibuat tentang produksi, manajemen rantai pasokan, dan pemasaran. Model prediksi diperlukan untuk memantau dinamika pasar masa lalu dan faktor eksternal, seperti kondisi ekonomi, tren populasi, dan perubahan nilai tukar. Hal ini karena industri otomotif dapat mempengaruhi kondisi pasar dan membuat permintaan pelanggan tidak pasti.

Oleh karena itu, tujuan utama penelitian ini adalah membangun model forecasting penjualan mobil yang lebih akurat dengan membandingkan tiga algoritma Machine Learning, yaitu LightGBM, Random Forest, dan XGBoost. Melalui pendekatan komparatif ini, penelitian ingin mengidentifikasi model yang paling optimal sehingga dapat digunakan sebagai dasar pengambilan keputusan bisnis yang lebih tepat sasaran, efisien, dan data-driven.

1. Data Understanding

Untuk memperoleh pemahaman data, struktur data penjualan mobil GAIKINDO dipelajari, yang mencakup jumlah penjualan dan variabel waktu untuk merek mobil seperti Toyota, Honda, Daihatsu, Mitsubishi, dan Suzuki. Penelitian juga melihat faktor luar, seperti nilai tukar rupiah, yang dapat mempengaruhi harga dan daya beli kendaraan. Pada titik ini, penyelidikan awal dilakukan, termasuk pengecekan nilai yang tidak ada, distribusi data, pola musiman, tren, dan korelasi antar fitur. Sebelum memasuki tahap pemodelan, proses ini memastikan bahwa karakteristik data benar-benar dipahami.



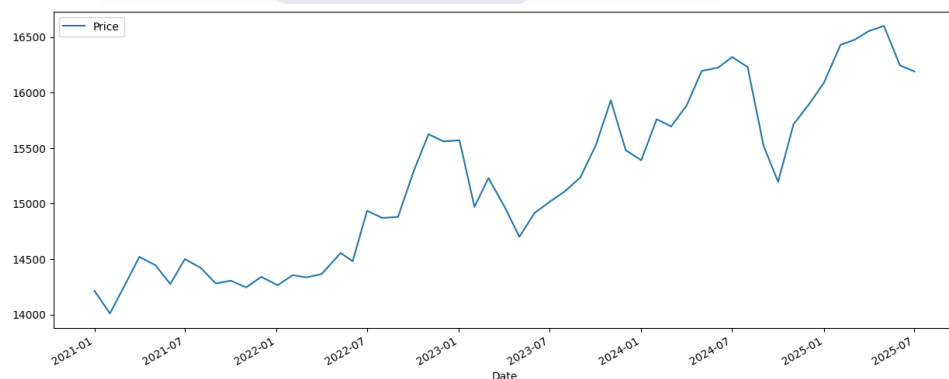
Gambar 3.4 Plot Time Series (utama)

Pada gambar 3.4, Tren penjualan bulanan dari lima merek mobil Daihatsu, Honda, Mitsubishi, Suzuki, dan Toyota mulai dari awal 2021 hingga 2025. Setiap garis mewakili merek tertentu, dan setiap garis memiliki warna yang berbeda untuk membantu Anda membandingkan bagaimana penjualan berubah sepanjang waktu.

Secara keseluruhan, Toyota tampaknya memiliki angka penjualan paling tinggi dibandingkan empat merek lainnya, dengan fluktuasi yang signifikan antara tahun 2021 dan 2023. Selain itu, Toyota terlihat mengalami beberapa puncak penjualan yang signifikan, menunjukkan periode penjualan yang sangat tinggi di beberapa bulan. Sementara itu, Daihatsu cenderung menjadi merek dengan penjualan menengah namun cukup stabil hingga 2023, sebelum mengalami penurunan pada 2024. Honda

dan Mitsubishi menunjukkan pola naik-turun yang moderat, dengan volume penjualan di bawah Daihatsu dan jauh di bawah Toyota. Suzuki terlihat memiliki angka penjualan paling rendah dari kelima merek, dengan tren yang relatif stabil namun cenderung menurun dari 2024 hingga 2025.

Selain itu, grafik menunjukkan pola penurunan kolektif untuk semua merek dari pertengahan hingga akhir 2023 dan memasuki 2024. Pola ini dapat menunjukkan adanya elemen eksternal yang berdampak pada industri secara keseluruhan, seperti perubahan ekonomi, peraturan, atau kondisi makro lainnya. Grafik ini memberikan gambaran makro tentang dinamika pasar otomotif Indonesia dalam beberapa tahun terakhir, termasuk periode pertumbuhan, periode fluktuasi tajam, dan fase penurunan yang dialami oleh semua merek secara bersamaan, karena memasuki tahun 2025, beberapa merek mulai menunjukkan pemulihan perlahan.

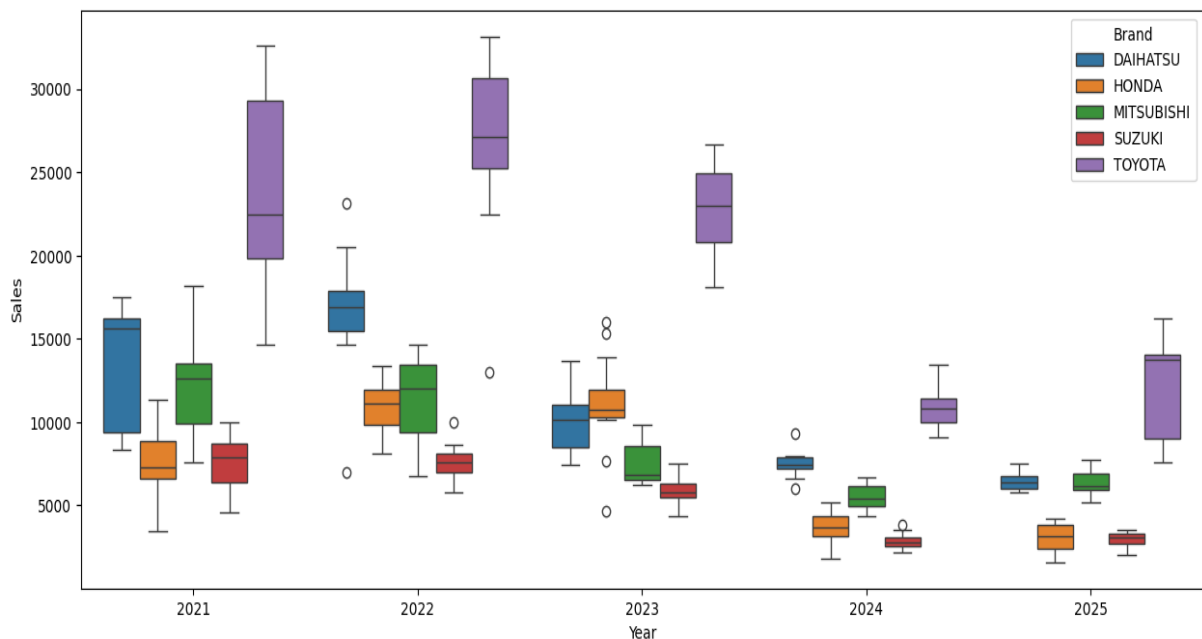


Gambar 3.5 Plot time-series (kurs)

Pada gambar 3.5 menunjukkan tren harga secara keseluruhan dari tahun 2021 hingga pertengahan 2025. Garis vertikal menunjukkan nilai harga, dan sumbu horizontal menunjukkan tanggal. Grafik biasanya menunjukkan tren kenaikan yang konsisten dari waktu ke waktu. Pada awal 2021, harga berada di kisaran 14.000–14.500, tetapi naik perlahan dari awal 2022 hingga pertengahan 2022. Pada pertengahan 2022 hingga 2023, tren kenaikan menjadi lebih jelas, dengan beberapa lonjakan besar terjadi, seperti pada akhir 2022 ketika harga melejit ke sekitar 15.600. Tren secara keseluruhan masih menunjukkan peningkatan, meskipun terjadi beberapa koreksi atau penurunan kecil seiring waktu, seperti di pertengahan tahun 2023.

Memulai tahun 2024, harga naik dengan cepat. Mereka mencapai puncaknya pada pertengahan tahun 2025, sekitar 16.700, tetapi setelah mencapai puncaknya, grafik menunjukkan sedikit penurunan menuju akhir rentang data, tetapi harga masih jauh lebih tinggi daripada pada awal periode.

Secara keseluruhan, gambar ini menunjukkan tren naik jangka panjang (upward trend) yang relatif stabil dengan beberapa variasi kecil. Pola ini dapat menunjukkan dinamika pasar seperti inflasi, peningkatan permintaan, perubahan biaya produksi, atau faktor lain yang secara konsisten mendorong harga ke level yang lebih tinggi setiap tahunnya.



Gambar 3.6 Boxplot Data Gaikindo

Pada gambar 3.6 menunjukkan boxplot penjualan mobil per tahun untuk periode 2021–2025 untuk lima merek mobil utama: Daihatsu, Honda, Mitsubishi, Suzuki, dan Toyota. Visualisasi ini juga menunjukkan data distribusi, variasi penjualan, dan pola perubahan pasar tahunan.

Boxplot biasanya menunjukkan median, rentang interkuartil (IQR), nilai

minimum-maksimum, dan outlier untuk masing-masing merek untuk masing-masing tahun. Dari sudut pandang interpretasi, grafik ini menunjukkan pola bahwa tahun 2021–2023 menunjukkan penjualan yang relatif tinggi untuk seluruh merek; ini terutama berlaku untuk Toyota, yang selama bertahun-tahun mencatat angka penjualan tertinggi. Selain itu, beberapa merek, seperti Daihatsu dan Mitsubishi, menunjukkan penyebaran nilai yang cukup besar pada tahun-tahun awal, yang menunjukkan fluktuasi permintaan yang lebih tinggi.

Memasuki tahun 2024, boxplot menunjukkan penurunan signifikan pada seluruh merek, ditandai dengan penurunan median dan mengecilnya IQR, yang mencerminkan fase pelemahan umum pasar otomotif. Namun, pada tahun 2025 terjadi sedikit pemulihan, terutama pada Toyota dan Mitsubishi, yang menunjukkan peningkatan median penjualan, meskipun belum kembali ke level awal.

Visualisasi ini tidak hanya menunjukkan distribusi penjualan setiap merek setiap tahun, tetapi juga menunjukkan tren pasar mobil di Indonesia. Pasar ini mengalami peningkatan besar pada 2021–2023, penurunan drastis di 2024, dan kemungkinan pemulihan pada 2025.

2. Data Preperation

Pada tahap persiapan data, diperlukan model prediktif yang mampu membaca kembali tren pasar sebelumnya serta berbagai faktor eksternal seperti kondisi ekonomi, pertumbuhan jumlah penduduk, dan perubahan nilai tukar. Hal ini menjadi penting karena pergerakan para pelaku industri dapat memengaruhi dinamika pasar dan membuat tingkat permintaan menjadi kurang stabil.

Penelitian ini bertujuan untuk meningkatkan akurasi model peramalan penjualan mobil dengan membandingkan tiga algoritma machine learning, yaitu LightGBM, Random Forest, dan XGBoost. Melalui perbandingan ini, diharapkan dapat ditemukan model yang paling sesuai dan dapat diandalkan sebagai dasar dalam pengambilan keputusan bisnis yang lebih cerdas, aman, dan efisien berbasis data.


```
df['waktu'] = pd.to_datetime(df['waktu'], format='%Y-%m-%d')

cols = ["DAIHATSU", "HONDA", "MITSUBISHI", "SUZUKI", "TOYOTA"]
df[cols] = df[cols].astype(int)

df.info()
```

Gambar 3.7 Mengubah format

Pada gambar 3.7, sebelum memasuki tahap analisis lanjutan dalam pemodelan deret waktu, dilakukan standardisasi tipe data, seperti yang ditunjukkan. Pada baris pertama, kolom waktu diubah dari representasi string menjadi tipe datetime64. `df['waktu'] = pd.to_datetime(df['waktu'], format='%Y-%m-%d')`. Konversi ini sangat penting karena struktur data bertipe waktu memungkinkan dataframe untuk melakukan berbagai operasi analitis berbasis waktu, seperti indeksasi waktu, pengeluaran lambat, dan resampling, serta meminimalkan kesalahan interpretasi tanggal data selama pemodelan. Langkah ini memastikan bahwa data temporal dianggap sebagai objek waktu, bukan teks.

Selanjutnya, perintah `df[cols] = df[cols]` digunakan untuk menyimpan variabel cols dengan daftar kolom yang menunjukkan jumlah penjualan dari masing-masing merek mobil. Setiap kolom diubah menjadi tipe integer dengan menggunakan `astype(int)`. Karena data penjualan tidak memerlukan representasi numerik desimal, penetapan tipe integer diperlukan. Selain itu, proses ini menghindari kemungkinan error ketika melakukan operasi matematis atau algoritmik pada tahap feature engineering dan pemodelan dengan memastikan konsistensi struktur data.

```
df_eksternal_1['Date'] = pd.to_datetime(df_eksternal_1['Date'], format='%m/%d/%Y')

df_eksternal_1 = df_eksternal_1.sort_values('Date')

df_eksternal_1_monthly = df_eksternal_1.groupby(df_eksternal_1['Date'].dt.to_period('M')).first().reset_index(drop=True)

print(df_eksternal_1_monthly)
```

Gambar 3.8 Mengubah Format Data kurs

Pada gambar 3.8, digunakan untuk membersihkan dan menyiapkan data kurs harian agar bisa digunakan sebagai variabel eksternal dalam analisis time series bulanan. Karena data utama (penjualan mobil Gaikindo) berada pada frekuensi bulanan, maka data kurs yang awalnya harian harus disesuaikan agar kedua dataset memiliki frekuensi waktu yang sama. Pertama, kode mengubah kolom Date menjadi format tanggal yang benar sehingga pandas dapat mengenali dan memprosesnya sebagai data waktu. Setelah itu, data disortir berdasarkan tanggal untuk memastikan urutan kronologis yang konsisten.

Tahap pentingnya adalah proses grouping berdasarkan bulan menggunakan `dt.to_period('M')`, kemudian mengambil nilai pertama dari setiap bulan. Proses ini digunakan untuk mengonversi data harian menjadi data bulanan dengan cara yang sederhana dan stabil. Hasil akhirnya adalah dataframe baru yang berisi satu nilai kurs per bulan, yang sudah selaras dengan data utama sehingga dapat langsung digabungkan dan dipakai sebagai fitur eksternal dalam pemodelan Machine Learning atau forecasting.

```
df_eksternal_2['Week'] = pd.to_datetime(df_eksternal_2['Week'], format='%Y-%m-%d')
df_eksternal_2["Month"] = df_eksternal_2["Week"].dt.to_period("M")

df_eksternal_2_monthly = df_eksternal_2.groupby("Month").sum(numeric_only=True).reset_index()
df_eksternal_2_monthly = df_eksternal_2_monthly.drop(index=0).reset_index(drop=True)

print(df_eksternal_2_monthly)
```

	Month	honda: (Indonesia)	daihatsu: (Indonesia)	toyota: (Indonesia)
0	2021-01	319	32	125
1	2021-02	257	28	100
2	2021-03	242	26	91
3	2021-04	245	28	106
4	2021-05	339	39	149
5	2021-06	241	27	97
6	2021-07	208	21	77
7	2021-08	269	29	102
8	2021-09	216	22	80
9	2021-10	275	30	107
10	2021-11	211	30	92
11	2021-12	223	26	89
12	2022-01	311	33	121

Gambar 3.9 Data Google Trend

Pada gambar 3.9, menyesuaikan dengan frekuensi data utama yang digunakan dalam analisis, ia mengubah data Google Trends yang sebelumnya dicatat dalam skala

mingguan menjadi agregasi bulanan, seperti yang ditunjukkan. Pertama, kolom "Minggu" harus diubah menjadi tipe datetime agar dapat diproses sebagai data deret waktu karena Google Trends memberikan nilai indeks pencarian per minggu. Setelah itu, setiap minggu dihitung ke jangka waktu bulan dengan menggunakan `dt.to_period(M)`. Transformasi ini penting karena pemodelan peramalan, analisis tren, dan korelasi sering membutuhkan keseragaman interval waktu antara variabel eksternal dan variabel target.

Untuk menghasilkan representasi total intensitas pencarian selama satu bulan, data yang memiliki label bulan kemudian dikelompokkan berdasarkan kolom "Bulan", dan nilai indeks pencarian mingguan kemudian dijumlahkan. Metode agregasi ini menunjukkan akumulasi minat masyarakat terhadap merek mobil tertentu selama satu bulan, yang memberikan gambaran yang lebih stabil dan informatif daripada data mingguan yang lebih fluktuatif.

Hasil akhirnya adalah dataframe bulanan yang menunjukkan volume pencarian Google Trends untuk masing-masing merek mobil setiap bulan. Dataset ini lebih mudah dianalisis untuk mengidentifikasi pola jangka panjang dan perubahan minat konsumen. Mereka juga lebih cocok untuk digunakan sebagai input eksternal dalam model peramalan penjualan atau analisis pasar lainnya.

```
# merge data utama setelah differencing + data eksternal 1 (nilai kurs)
df_merge_diff = df_diff.merge(df_eksternal_1_monthly, on="waktu", how="left")

# merge hasil sebelumnya + eksternal 2 (google trends)
df_merge_diff = df_merge_diff.merge(df_eksternal_2_monthly, on="waktu", how="left")

# merge hasil sebelumnya + eksternal 3 (inflasi transportasi)
df_merge_diff = df_merge_diff.merge(df_eksternal_3, on="waktu", how="left")

df_merge_diff.head()
```

Gambar 3.10 Merge Semua Data

Pada gambar 3.10, data penjualan mobil digabungkan dengan tiga sumber data eksternal nilai kurs, tren Google, dan inflasi sektor transportasi. Metode *merge* kiri

yang didasarkan pada kolom waktu, digunakan secara bertahap untuk menggabungkan semua data utama sekaligus mendapatkan variabel tambahan dari setiap sumber eksternal. Pada tahap pertama, data utama dan data kurs bulanan digabungkan untuk menunjukkan variabel ekonomi makro yang dapat memengaruhi penjualan kendaraan. Selanjutnya, hasil penggabungan ditambahkan ke data Google Trends, yang menunjukkan minat pencarian masyarakat terhadap merek-merek mobil, yang menunjukkan tingkat perhatian atau potensi permintaan konsumen. Pada tahap terakhir, data inflasi transportasi digabungkan untuk menunjukkan variabel apa yang dapat memengaruhi penjualan kendaraan.

Merge terakhir menghasilkan tabel terpadu yang berisi variabel penjualan dari lima merek mobil (DAIHATSU, HONDA, MITSUBISHI, SUZUKI, dan TOYOTA). Tabel ini diikuti oleh variabel eksternal seperti Nilai_Kurs, Inflasi_Transportasi, dan berbagai indeks Google Trends (GT_Honda, GT_Daihatsu, GT_Toyota, GT_Suzuki, GT_Mitsubishi). Karena setiap baris menunjukkan satu titik waktu bulanan, analisis lanjutan dapat dilakukan secara bertahap untuk mengidentifikasi hubungan antara penjualan kendaraan dan indikator ekonomi serta minat pencarian digital. Struktur data ini memungkinkan model prediksi menggunakan berbagai faktor eksternal, yang dianggap dapat meningkatkan akurasi prediksi penjualan dengan mempelajari pola lebih lanjut.

3. pembuatan fitur lag

```
# Lag features
for lag in lags:
    df[f'{brand}_lag{lag}'] = df[brand].shift(lag).bfill()
```

Gambar 3.11 Fitur Lag

Pada Gambar 3.11 terlihat bagaimana data penjualan dari beberapa periode sebelumnya digunakan dalam proses pembentukan fitur untuk model prediksi. Informasi seperti penjualan satu bulan, dua bulan, hingga tiga bulan sebelumnya dimasukkan sebagai variabel lag. Teknik ini penting karena penjualan di satu periode

biasanya dipengaruhi oleh kinerja di periode-periode sebelumnya, terutama pada data time series yang memang memiliki hubungan antarwaktu.

Tujuan penggunaan fitur lag adalah agar model dapat menangkap pola-pola yang pernah muncul di masa lalu misalnya tren naik-turun yang berulang, perubahan musiman, atau efek momentum penjualan dari waktu ke waktu. Dengan cara ini, model bisa memahami konteks historis yang menjadi dasar dalam menghasilkan prediksi yang lebih akurat. Dengan cara ini, model dapat membaca dinamika penjualan masa lalu sebagai acuan dalam memprediksi nilai di masa mendatang. Dengan mengidentifikasi pola ini, model dapat membuat prediksi yang lebih stabil dan realistis karena mengandalkan kondisi saat ini dan dinamika yang terjadi di masa lalu.

Kode penelitian menggunakan fungsi shift, atau lag, untuk mengubah nilai data ke bawah selama periode lag yang diinginkan. Nilai kosong (NaN) pada baris pertama diatasi dengan metode backfill (bfill()) setelah proses penggeseran. Metode ini mengisi nilai kosong dengan nilai valid berikutnya agar tidak terjadi error pada proses pelatihan model. Karena sebagian besar algoritma pembelajaran mesin tidak dapat memproses data dengan nilai hilang, penanganan ini sangat penting.

4. rolling window

```
def create_features(df, brands, rolling_windows=[3], lags=[1,2,3]):  
    df = df.copy()  
    df['month'] = df['waktu'].dt.month  
    df['month_sin'] = np.sin(2 * np.pi * df['month'] / 12)  
    df['month_cos'] = np.cos(2 * np.pi * df['month'] / 12)
```

Gambar 3.12 Coding Rolling Window

Pada gambar 3.12 rata-rata bergerak, atau moving average, adalah rata-rata jumlah penjualan selama tiga bulan terakhir. Perhitungan ini dilakukan menggunakan jendela bergerak (rolling window). Metode ini memudahkan untuk melihat pola tren yang sebenarnya, tetapi perubahan besar dalam nilai dari hari ke hari atau bulan ke bulan tidak mempengaruhi hasilnya. Dengan menghitung rata-rata bergerak pada setiap titik waktu, model dapat menangkap perubahan seiring waktu yang tidak terlihat jelas saat melihat nilai mentah saja.

Metode ini sangat berguna untuk analisis deret waktu karena dapat membantu mengidentifikasi perubahan dalam pola musiman, tren jangka pendek, dan tanda-tanda anomali yang mungkin terjadi. Selain itu, jendela bergerak membuat masukan untuk model pembelajaran mesin lebih stabil, yang sangat penting untuk algoritma yang sensitif terhadap noise. Hal ini membuat hasil prediksi lebih akurat dan konsisten. Jendela bergerak merupakan bagian penting dari proses prapemrosesan data karena memberikan informasi waktu yang lebih rinci dan membantu kita membuat prediksi yang lebih.

5. seasonal decomposition

Seasonal decomposition dilakukan dengan memisahkan data seri waktu menjadi tiga komponen utama: trend, seasonality, dan residual. Sebelum memulai proses modeling, langkah ini dilakukan untuk mendapatkan pemahaman tentang pola dasar yang ada dalam data. Pertama, data diatur ulang dengan menggunakan `df.set_index('waktu')` untuk menjadikan kolom waktu sebagai indeks.

```
y = df['DAIHATSU']
y = y.asfreq('MS')

decomposition = sm.tsa.seasonal_decompose(y, model='additive', period=12)
fig = decomposition.plot()
plt.show()
```

Gambar 3.13 Seasonal Decomposition

Pada gambar 3.12, library statistik dapat menyebut dataset sebagai deret waktu. Untuk memungkinkan analisis musiman dilakukan dengan benar, contohnya salah satu variabel penjualan HONDA, dipilih dan frekuensinya diubah menjadi bulanan menggunakan `asfreq`. Kemudian, perintah `sm.tsa.seasonal_decompose` memecah data menggunakan model additive dengan periode dua belas bulan, yang berarti pola musiman dianggap berulang setiap dua belas bulan.

Untuk menunjukkan bagaimana tren jangka panjang, pola musiman tahunan,

dan komponen acak (suara) terbentuk dalam penjualan Honda, hasil `decomposition.plot()` digunakan. Gambaran ini sangat penting untuk mengevaluasi stabilitas pola musiman, menemukan anomali, dan memberikan gambaran awal apakah data perlu diubah sebelum diproses dalam model pembelajaran mesin.

6. Scaling

```
from sklearn.preprocessing import StandardScaler

def scale_features(df, exclude_cols):
    feature_cols = [c for c in df.columns if c not in exclude_cols]
    scaler = StandardScaler()
    df_scaled = df.copy()
    df_scaled[feature_cols] = scaler.fit_transform(df_scaled[feature_cols])
    return df_scaled, scaler
```

Gambar 3.14 Scaling data

Pada gambar 3.13, `StandardScaler()` digunakan untuk menghilangkan nilai mean dan membagi data dengan standar deviasi sehingga menghasilkan distribusi dengan mean = 0 dan standar deviasi = 1. Dalam praktiknya, sering kali ada beberapa kolom yang tidak perlu diskalakan, misalnya kolom identifier atau kolom binary/one-hot, sehingga dibuat daftar `exclude_cols` untuk mengecualikan kolom tersebut. Ketika kita menjalankan `scaler.fit_transform(df[cols])`, scaler akan mempelajari parameter statistik (mean dan standar deviasi) dari seluruh dataset, lalu mentransformasi seluruh data berdasarkan nilai tersebut.

7. Pembuatan fitur interaksi dan rolling forecast

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, mean_absolute_percentage_error
import numpy as np

def rolling_forecast_log(df, target, model, initial_train_size):
    X = df.drop(columns=[target]).values
    y = np.log1p(df[target].values)

    y_true, y_pred = [], []
    for i in range(initial_train_size, len(df)):
        X_train, y_train = X[:i], y[:i]
        X_test = X[i:i + 1]
        model.fit(X_train, y_train)
        pred = np.exp1(model.predict(X_test)[0])
        y_true.append(np.exp1(y[i]))
        y_pred.append(pred)

    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return {
        'y_true': y_true,
        'y_pred': y_pred,
        'MAPE (%)': mean_absolute_percentage_error(y_true, y_pred) * 100
    }
```

Gambar 3.15 coding rolling forecast

Pada gambar 3.14 adalah proses dimulai dari ukuran awal data latih (initial train size) hingga akhir dataset, di mana pada setiap iterasi data sebelum indeks tertentu digunakan sebagai data latih, sedangkan baris pada indeks tersebut dijadikan data uji satu langkah ke depan. Dengan demikian, model hanya memanfaatkan informasi masa lalu untuk melakukan prediksi sehingga tidak terjadi lookahead bias. Selanjutnya, fitur dan target dipisahkan melalui pemisahan variabel independen (X_{train}) dan variabel dependen (y_{train}). Model kemudian dilatih ulang menggunakan data latih pada setiap iterasi (re-train every step), meskipun pendekatan ini relatif mahal secara komputasi. Tahap ini menuntut ketepatan konfigurasi parameter model serta pemilihan hyperparameter yang konsisten

8. Modeling

Dalam penelitian ini, tiga model ML berbasis ensemble LightGBM, XGBoost, dan Random Forest dikonfigurasi dengan parameter yang setara: `n_estimator` jumlah pohon=700, kedalaman maksimum maksimal=7, tingkat pengambilan sampel `subsample`=0.8, dan proporsi fitur `colsample_bytree`=0.8. Langkah ini dilakukan untuk memastikan bahwa proses perbandingan kinerja masing-masing model dilakukan dengan adil dan tanpa terpengaruh oleh perbedaan pengaturan yang dapat menyebabkan bias. Fokus penelitian ini bukan untuk mengevaluasi berbagai variasi hyperparameter, tetapi untuk melihat seberapa baik masing-masing algoritma mampu memprediksi permintaan mobil berdasarkan data deret waktu yang sama. Oleh karena itu, penyesuaian ini penting.

```
from xgboost import XGBRegressor
from copy import deepcopy

# --- XGBoost model ---
xgb_model = XGBRegressor(
    n_estimators=700, learning_rate=0.03, max_depth=7,
    subsample=0.8, colsample_bytree=0.8, objective='reg:squarederror',
    random_state=42
)

metrics_xgb = {}
for brand in brands:
    metrics_xgb[brand] = rolling_forecast_log(df_feat_scaled, brand, deepcopy(xgb_model), initial_train_size)
```

Gambar 3.16 Modeling XGBoost

Pada gambar 3.15 XGBoost, dibangun menggunakan parameter yang sama, seperti `n_estimators`=700, `learning_rate`=0.03, dan `max_depth`=7. Penyamaan parameter ini memastikan bahwa perbandingan kinerja antara XGBoost dan LightGBM bukan disebabkan oleh tuning yang berbeda. Dengan menggunakan `deepcopy`, proses pelatihan setiap merek berlangsung independen. XGBoost kemudian menjalani proses yang sama melalui `rolling_forecast_log`, dengan hasil evaluasi disimpan di dalam `metrics_xgb`. Dengan karakteristiknya yang kuat dalam mengatasi data non-linear, XGBoost menjadi kandidat penting dalam komparasi ini.

```

from sklearn.ensemble import RandomForestRegressor
from copy import deepcopy

# --- RandomForest model ---
rf_model = RandomForestRegressor(
    n_estimators=700, max_depth=7, min_samples_split=2, min_samples_leaf=1, random_state=42
)

metrics_rf = {}
for brand in brands:
    metrics_rf[brand] = rolling_forecast_log(df_feat_scaled, brand, deepcopy(rf_model), initial_train_size)

```

Gambar 3.17 Modeling Random Forest

Pada gambar 3.17, Random Forest dibangun dengan parameter yang disamakan, seperti `n_estimators=700` dan `max_depth=7`, untuk menjaga fairness komparasi. Model ini juga di-deepcopy untuk setiap merek sebelum diproses melalui fungsi `rolling_forecast_log`. Hasil evaluasinya kemudian disimpan pada `metrics_rf`. Random Forest berfungsi sebagai pembanding yang baik untuk melihat perbedaan kinerja antara metode boosting dan metode bagging dalam konteks forecasting time series.

```

1 from lightgbm import LGBMRegressor
2 from copy import deepcopy
3
4 # --- LightGBM model ---
5 lgb_model = LGBMRegressor(
6     n_estimators=700, learning_rate=0.03, max_depth=7,
7     num_leaves=40, subsample=0.8, colsample_bytree=0.8,
8     objective='regression', random_state=42, verbose=-1
9 )
10
11 metrics_lgb = {}
12 for brand in brands:
13     metrics_lgb[brand] = rolling_forecast_log(df_feat_scaled, brand, deepcopy(lgb_model), initial_train_size)

```

Gambar 3.18 Modeling LightGBM

Pada gambar 3.18, proses pemodelan menggunakan algoritma LightGBM dilakukan untuk seluruh mobil secara terstruktur dan konsisten. Pertama, sebuah model dasar LightGBM (`lgb_model`) didefinisikan dengan parameter yang telah ditentukan secara manual, seperti `n_estimators`, `learning_rate`, `max_depth`, dan `num_leaves`. Parameter-parameter ini mengontrol kompleksitas model, kecepatan pembelajaran, serta kapasitas model dalam menangkap pola non-linear dari data penjualan mobil. Selain itu, sampling acak terhadap baris dan fitur pada setiap iterasi pembentukan pohon keputusan digunakan untuk mencegah overfitting dengan `subsample` dan `colsample_bytree`. Random Forest menawarkan baseline yang kuat

melalui teknik bagging yang stabil dan tahan terhadap overfitting, sedangkan LightGBM dan XGBoost menampilkan pendekatan gradient boosting yang lebih agresif untuk menangkap pola non-linear. Pendekatan multi-model ini sangat penting untuk memastikan bahwa rekomendasi model penelitian benar-benar reliabel, objektif, dan dapat diterapkan dalam skenario prediksi dunia nyata.

3.1.1.7 Penyusunan Artikel Ilmiah

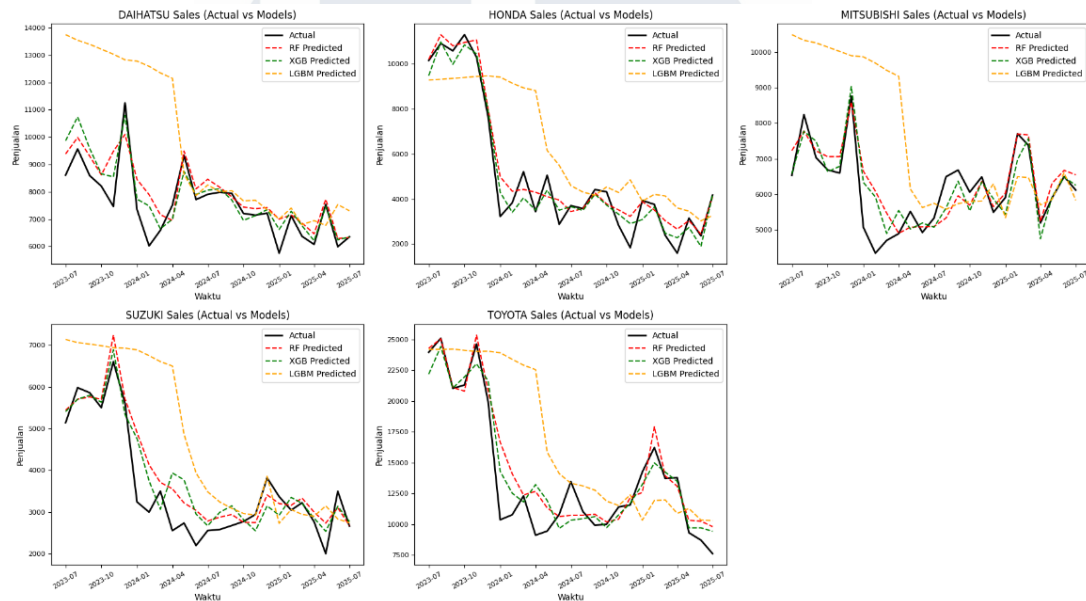
Proses evaluasi model dilakukan selama tahap penyusunan artikel ilmiah untuk menilai performa tiga algoritma yang digunakan dalam penelitian: XGBoost, Random Forest, dan LightGBM. Proses ini dilakukan dengan menggunakan metrik Mean Absolute Percentage Error (MAPE) untuk mengevaluasi tingkat akurasi hasil ramalan terhadap data aktual.

Tabel 2.2 Evaluasi MAPE

Metrik/Merek	MAPE (%)		
	RF	XGB	LGBM
Daihatsu	7.78	5.57	30.35
Honda	16.25	13.32	52.58
Mitsubishi	7.86	7.17	32.58
Suzuki	13.15	14.16	39.18
Toyota	11.54	10.52	34.96
Rata-rata	11.32	10.15	37.89

Pada tabel 3.2. hasil pengujian menunjukkan bahwa XGBoost memiliki performa terbaik dengan nilai MAPE sebesar 10,15%, menunjukkan bahwa model ini mampu menangkap pola non-linear dan variasi pada data penjualan mobil dengan lebih konsisten. RandomForest memiliki nilai MAPE kedua terbaik sebesar 11.32%. Namun, LightGBM mendapatkan MAPE terburuk dengan sebesar 37,89%, menunjukkan kinerja yang lebih buruk daripada dua model lainnya. Hasil ini menunjukkan bahwa LightGBM kurang mampu beradaptasi dengan struktur data yang digunakan. Ini mungkin karena sensitivitasnya terhadap distribusi fitur, ukuran dataset

yang tidak terlalu besar, dan kompleksitas fitur yang dihasilkan oleh proses feature engineering. Setelah itu, perbedaan performa dari ketiga model ini dianalisis dan dijelaskan secara mendalam dalam artikel ilmiah; ini menjadi dasar untuk membuat kesimpulan dan kontribusi penelitian.



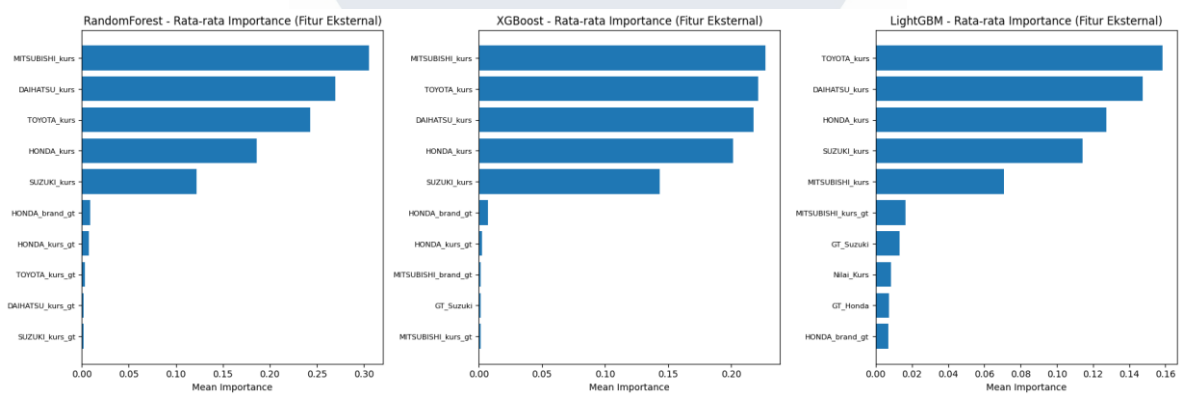
Gambar 3.19 Predicted vs Actual Semua model

Pada gambar 3.19 membandingkan data penjualan asli dengan hasil prediksi dari tiga model machine learning: Random Forest (RF), XGBoost (XGB), dan LightGBM (LGBM). Kelima merek mobil yang dianalisis adalah Daihatsu, Honda, Mitsubishi, Suzuki, dan Toyota. Garis berwarna hitam menunjukkan data penjualan asli, sementara garis putus-putus menunjukkan hasil prediksi dari masing-masing model.

Secara umum, penjualan seluruh merek mengalami penurunan dari waktu ke waktu, meskipun masing-masing merek tetap menunjukkan pola naik dan turun yang berbeda. Model RF dan XGB mampu mengikuti pola penjualan asli dengan cukup baik. Hal ini terlihat dari garis prediksi yang posisinya lebih dekat dengan garis data sebenarnya di hampir seluruh periode. Kedua model ini juga cukup konsisten dalam menangkap perubahan tren serta fluktuasi penjualan setiap bulan.

Di sisi lain, model LGBM cenderung memprediksi angka yang lebih tinggi, khususnya pada awal periode analisis. Pola ini menunjukkan bahwa LGBM memberikan estimasi yang terlalu optimistis dibandingkan data riil. Karena kecenderungan tersebut muncul pada semua merek, hal ini menandakan bahwa LGBM masih memerlukan penyesuaian tambahan agar dapat menangani tren penurunan yang tajam secara lebih akurat.

Grafik ini menunjukkan bahwa RF dan XGB menghasilkan prediksi yang lebih akurat dan stabil dibandingkan LGBM, sehingga keduanya lebih cocok digunakan untuk pemodelan penjualan mobil berbasis deret waktu.



Gambar 3.20 Feature Importance

Pada gambar 3.19 adalah untuk memahami peran masing-masing fitur dalam memengaruhi hasil prediksi penjualan mobil yang menunjukkan visualisasi pentingnya fitur variabel eksternal pada tiga model yang digunakan Random Forest, XGBoost, dan LightGBM. Ketiga grafik menunjukkan rata-rata pentingnya fitur (mean importance) terhadap performa model, yang menunjukkan fitur mana yang paling berpengaruh dalam membentuk output prediksi.

Variabel kurs untuk merek Mitsubishi, Daihatsu, dan Toyota adalah fitur eksternal yang paling penting pada model Random Forest. Hal ini menunjukkan bahwa perubahan nilai tukar memiliki dampak yang signifikan terhadap jumlah barang yang dijual oleh ketiga merek tersebut. Meskipun kontribusinya lebih sedikit, Honda dan

Suzuki juga sangat berpengaruh terhadap kurs. Namun, variabel Google Trends (brand_gt) memberikan kontribusi yang lebih kecil, menunjukkan bahwa hasil prediksi model ini tidak terlalu dipengaruhi oleh perilaku pencarian internet.

Pola yang dihasilkan pada model XGBoost hampir identik. Dengan kurs Mitsubishi, Toyota, dan Daihatsu berada di tiga teratas, variabel kurs kembali menjadi faktor yang paling penting. Kurs Honda dan Suzuki juga memiliki dampak yang signifikan. Fitur pencarian Google Trends masih memberikan kontribusi yang sangat kecil, seperti yang ditunjukkan dalam model sebelumnya. Dengan konsistensi ini, XGBoost menunjukkan bahwa ia juga mempertimbangkan kondisi makroekonomi, dengan nilai tukar ditunjukkan sebagai faktor yang paling signifikan yang memengaruhi penjualan kendaraan.

Namun, dominasi variabel kurs menjadi lebih jelas pada LightGBM. Kurs Toyota, Daihatsu, Honda, dan Suzuki berada di posisi teratas dengan nilai signifikan. Dibandingkan dengan dua model sebelumnya, model ini menunjukkan kontribusi yang lebih baik untuk beberapa fitur Google Trends, tetapi kontribusinya masih agak kecil. Selain itu, variabel eksternal seperti inflasi dan penjualan makro lainnya (seperti GF_Suzuki atau GF_Honda) muncul, tetapi kontribusinya jauh lebih kecil daripada variabel kurs.

Secara keseluruhan, ketiga model menunjukkan pola yang sama: elemen eksternal yang paling berpengaruh terhadap prediksi penjualan seluruh merek otomotif adalah nilai tukar, sementara data pencarian Google Trends dan komponen eksternal lainnya memiliki dampak yang lebih kecil. Menurut temuan ini, permintaan kendaraan di pasar otomotif Indonesia dipengaruhi secara signifikan oleh faktor ekonomi makro, khususnya fluktuasi kurs.

3.1.1.8 Pengumpulan Full Paper dan Registrasi Ulang

Setelah mereka menyelesaikan seluruh makalah, peserta menggunakan sistem yang disiapkan oleh panitia untuk mengunggah manuskrip mereka. Pada titik ini,

proses registrasi ulang juga berlangsung. Tujuan dari proses ini adalah untuk memastikan bahwa semua dokumen yang dikirim memenuhi standar kompetisi untuk gaya penulisan, kelengkapan berkas pendukung, dan kepatuhan terhadap pedoman teknis yang diberikan. Langkah ini sangat penting bagi pihak administrasi karena akan membantu mereka memutuskan apakah peserta siap untuk melanjutkan ke tahap penilaian akhir.

Panitia juga biasanya melakukan pemeriksaan tambahan untuk memastikan bahwa manuskrip yang dikirim memenuhi standar akademik dari sebuah kompetisi nasional. Ini termasuk memastikan bahwa karya tersebut asli, informasi peserta lengkap, dan penulisan mengikuti aturan penulisan ilmiah. Langkah ini juga memastikan bahwa semua informasi teknis dan komunikasi mengenai babak final dikirimkan kepada orang yang tepat.

3.1.1.9 Pengumuman 10 Besar Finalis

Selanjutnya, panitia mengumumkan sepuluh finalis terbaik yang berhak melanjutkan ke babak presentasi. Kelulusan ini didasarkan pada penilaian kualitas kertas lengkap. Kualitas ini mencakup ketepatan metodologi, kedalaman analisis, kontribusi ilmiah, dan inovasi penelitian. Dalam proses seleksi, hal-hal seperti kejelasan alur penulisan, relevansi topik dengan tema kompetisi, dan ketepatan penggunaan data juga dipertimbangkan.

Proses penyaringan akhir dilakukan untuk menentukan peserta yang memiliki potensi akademik terbaik dan layak untuk mempresentasikan hasil penelitiannya kepada juri. Oleh karena itu, daftar sepuluh finalis terbaik menunjukkan kualitas riset terbaik serta kesiapan peserta untuk mempresentasikan argumen ilmiahnya pada presentasi nasional.

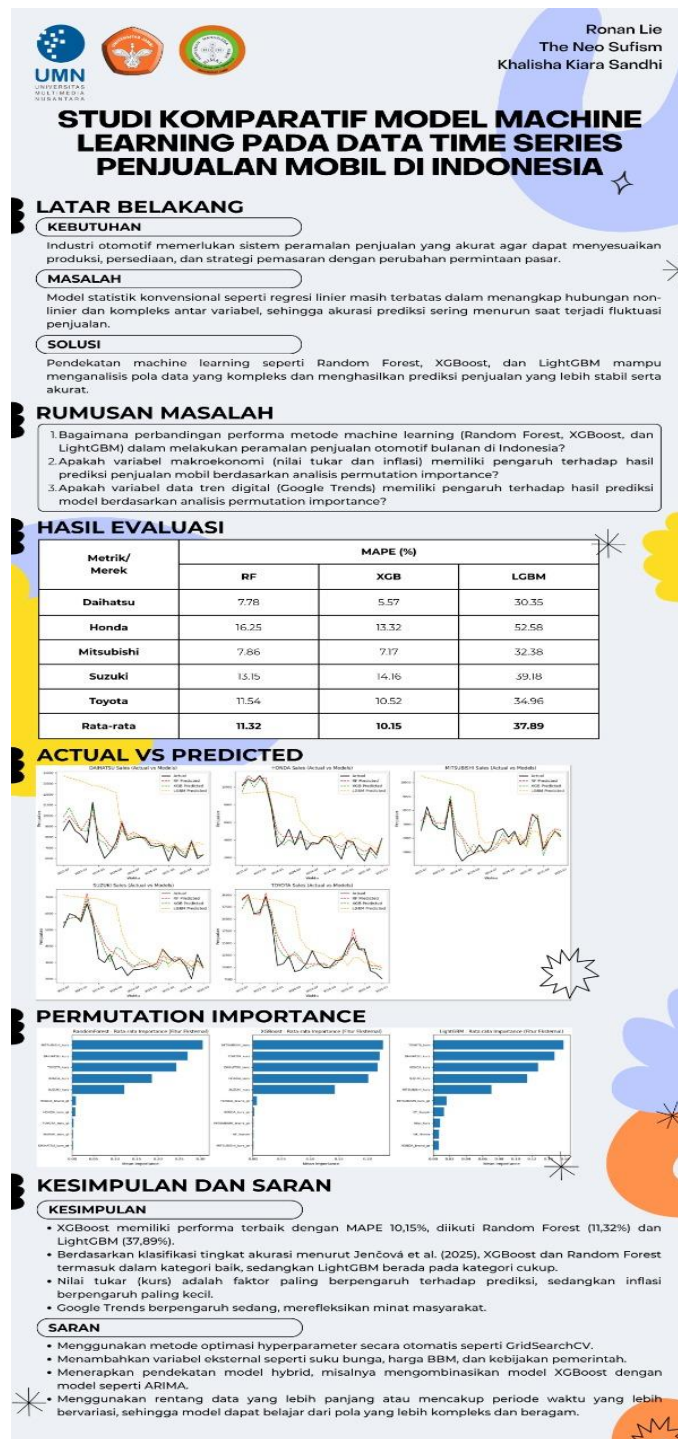
3.1.1.10 Technical Meeting dan Mengumpul Poster

Peserta yang lolos akan mengikuti pertemuan teknis bersama komite. Pada tahap ini, panitia memberikan penjelasan rinci terkait mekanisme presentasi, format penilaian, durasi presentasi, serta alur sesi tanya jawab pada babak final. Informasi ini diberikan agar setiap finalis memiliki gambaran yang jelas mengenai proses yang akan mereka lalui.

Selain itu, finalis mendapatkan arahan mengenai persyaratan teknis seperti format slide presentasi, aturan penggunaan media pendukung, serta prosedur registrasi ulang pada hari pelaksanaan. Panitia juga menekankan aturan kedisiplinan waktu agar seluruh rangkaian acara dapat berjalan dengan tertib. Hal-hal ini diberikan secara detail agar peserta tidak salah dalam mempersiapkan kebutuhan final.

Untuk menghindari miskomunikasi, panitia menyediakan sesi tanya jawab agar finalis dapat mengonfirmasi hal-hal teknis yang masih belum dipahami. Sesi ini memberi ruang bagi peserta untuk memastikan bahwa seluruh persyaratan dan prosedur telah dipahami dengan benar. Dengan demikian, proses final dapat berlangsung tanpa hambatan akibat kesalahan informasi.

Pertemuan teknis menjadi bagian penting untuk menyamakan persepsi seluruh finalis. Tujuannya adalah agar setiap peserta memahami mekanisme kompetisi secara jelas dan tidak ada pihak yang dirugikan akibat ketidaktahuan teknis. Melalui mekanisme ini, panitia memastikan sesi final berlangsung adil, profesional, dan tertib.



Gambar 3.21 Poster full-paper

Pada gambar 3.21, peserta harus mengumpulkan poster ilmiah yang berisi ringkasan dari makalah yang telah mereka tulis. Poster tersebut kemudian dicetak dan dipajang di area penilaian. Banner ini tidak hanya menampilkan latar belakang, metodologi, hasil analisis, dan kesimpulan, tetapi juga merupakan media visual utama

yang akan dinilai secara langsung oleh dewan juri.

Pada tahap penilaian poster, juri menilai kualitas desain, kerapian, kemampuan menyampaikan informasi secara ringkas, dan konsistensi antara poster dan isi kertas. Dengan komponen ini, finalis tidak hanya dinilai dari kemampuan presentasi mereka secara lisan, tetapi juga dari kualitas visual penelitian secara visual dan profesional.

3.1.1.11 Presentasi

Presentasi akhir dilakukan pada hadapan dewan juri yang terdiri dari praktisi dan akademisi di bidang yang relevan, seperti yang ditunjukkan , peserta mempresentasikan seluruh hasil kertas. Latar belakang penelitian, rumusan masalah, metodologi yang digunakan, proses analisis data, hasil prediksi, dan dampak penelitian terhadap kemajuan akademik dan bisnis dibahas dalam presentasi ini.



Gambar 3.22 Presentasi

Pada gambar 3.22 peserta diminta untuk memberikan penjelasan ringkas

namun mendalam tentang hasil. Penjelasan ini harus menggarisbawahi kontribusi utama penelitian, alasan mengapa metode analisis dipilih, dan relevansi hasil dengan situasi dunia nyata. Agar juri dapat menilai kualitas analisis secara menyeluruh, penyampaian data dan visualisasi pendukung harus dilakukan secara jelas, terstruktur, dan didukung oleh argumen metodologis yang konsisten.

Sebagai bentuk pertahanan argumen ilmiah, peserta terlibat dalam sesi tanya jawab selain memberikan materi. Juri menguji pemahaman peserta tentang metode penelitian pada sesi ini. Mereka juga menguji validitas proses analisis, legitimasi model yang dipilih, dan konsistensi antara data, hasil, dan kesimpulan penelitian. Pertanyaan dapat mencakup elemen teknis, konseptual, maupun implikatif, sehingga peserta harus menunjukkan kemampuan berpikir kritis, logis, dan respons yang berbasis data. Nilai penilaian juga meningkat jika penjelasan yang tepat tentang keterbatasan penelitian dan potensi pengembangannya di masa depan diberikan.

Setiap peserta diberi waktu tidak lebih dari sepuluh menit untuk menyampaikan hasil penelitiannya secara mendalam dan sistematis. Setelah itu, sesi dilanjutkan dengan tanya jawab selama lima belas menit yang dilakukan oleh tiga orang juri yang terdiri dari pakar dan dosen dalam bidang yang dibahas. Para juri akan memeriksa ketajaman analisis, validitas metode, dan relevansi hasil penelitian. Terdapat sepuluh finalis dari berbagai kota di Indonesia yang bersaing untuk tempat terbaik. Ini membuat sesi presentasi menjadi ajang kompetisi di mana setiap peserta harus memiliki kemampuan akademik dan kemampuan komunikasi ilmiah yang baik.

Dalam tahap akhir ini, tidak hanya dinilai kualitas teknis penelitian, tetapi juga dinilai kemampuan komunikasi ilmiah, profesionalisme dalam presentasi, dan kesiapan peserta untuk mempertahankan karya akademik mereka. Oleh karena itu, fase ini berfungsi sebagai proses seleksi komprehensif untuk memilih peserta yang memiliki kombinasi terbaik dari ketajaman analisis, ketepatan metodologi, inovasi penelitian, dan kemampuan untuk menyampaikan ide-ide tersebut secara efektif di ruang akademik. Tahap akhir ini tidak hanya berfungsi sebagai ajang penilaian, tetapi juga berfungsi sebagai ruang pembelajaran yang membantu peserta mendapatkan

pengalaman yang lebih baik dalam presentasi karya ilmiah secara formal dan kompetitif.

3.1.1.12 Pengumuman Pemenang

Pengumuman pemenang adalah langkah terakhir. Dalam kompetisi ini, peserta diberi penghargaan untuk masing-masing kategori. Kategori terbaik adalah "Presentasi Terbaik", yang diberikan kepada peserta karena mereka dapat menyampaikan materi dengan baik, memiliki struktur presentasi yang jelas, dan memiliki argumen ilmiah yang kuat. Penghargaan ini merupakan cara untuk mengakui kualitas komunikasi ilmiah dan pemahaman mendalam tentang hasil penelitian.



Gambar 3.23 Pengumuman Best Presentation

Pada gambar 3.23, Dua kategori penghargaan yaitu Best Poster dan Best Presentation. Tim yang berpartisipasi dalam kompetisi tersebut berhasil meraih penghargaan Best Presentation, mengungguli finalis lainnya dari berbagai kota. Kami menerima sertifikat penghargaan dan hadiah uang tunai sebesar 150 ribu rupiah sebagai tanda penghargaan. Penghargaan ini menunjukkan bahwa penyampaian materi, penjelasan, dan penguasaan data kami paling kuat dan meyakinkan.



Gambar 3.24 Penerimaan Sertifikat

Pada gambar 3.24 adalah upacara penyerahan sertifikat, yang berlangsung sehari setelah sesi presentasi. Hari terakhir kompetisi di Universitas Jambi mencakup upacara penghargaan ini sebagai bagian dari upacara penutupan. Setelah melalui proses seleksi, penilaian poster, dan presentasi di hadapan tiga juri, tim kami dinobatkan sebagai pemenang penghargaan Presentasi Terbaik. Penghargaan ini diberikan langsung sebelum adanya acara talkshow, bersamaan dengan pengumuman seluruh pemenang dari berbagai kategori dan juga pemenang keseluruhan juara. Selain sertifikat resmi, tim kami juga mendapatkan hadiah uang sebesar 150 ribu rupiah sebagai bentuk apresiasi. Momen ini menandai penutup yang membanggakan dari seluruh rangkaian kegiatan yang diikuti para finalis dari berbagai kota

3.1.1.13 *Post-Competition*

Post competition menunjukkan versi yang telah diperbarui sebagai peningkatan kode terhadap metode ramalan sebelumnya. Versi ini mengutamakan optimisasi parameter menggunakan Optuna, sehingga konfigurasi model setiap merek benar-benar disesuaikan dengan karakteristik datanya. Sehingga model tidak terlalu kompleks, fitur yang digunakan cukup sederhana dan hanya mencakup variabel waktu seperti bulan, kuartal, dan tahun. Untuk mensimulasikan kondisi prediksi dunia nyata, model dievaluasi menggunakan metode walk-forward forecasting setelah parameter terbaik diperoleh melalui proses hyperparameter tuning. Dalam metode ini, model dilatih menggunakan data historis hingga periode tertentu dan kemudian memprediksi kondisi secara bertahap satu per satu. Metode yang lebih sistematis ini menghasilkan model yang stabil dan menunjukkan kinerja LightGBM yang lebih realistis.



```

def walk_forward(df, target_col, best_params):

    df = df.copy()
    predictions = []
    actuals = []

    for i in range(6, len(df)):
        train = df.iloc[:i]
        test = df.iloc[i:i+1]

        X_train = train.drop(columns=[target_col, "waktu"])
        y_train = train[target_col]
        X_test = test.drop(columns=[target_col, "waktu"])
        y_test = test[target_col].values[0]

        model = lgb.LGBMRegressor(**best_params)
        model.fit(X_train, y_train)

        pred = model.predict(X_test)[0]
        predictions.append(pred)
        actuals.append(y_test)

    mape = mean_absolute_percentage_error(actuals, predictions) * 100
    return predictions, actuals, mape

brand_cols = ["DAIHATSU", "HONDA", "MITSUBISHI", "SUZUKI", "TOYOTA"]
results = {}

for brand in brand_cols:
    print(f"\n🔵 Processing {brand} ...")

    df_brand = build_features(df, brand)

    study = optuna.create_study(direction="minimize")
    study.optimize(lambda trial: objective(trial, df_brand, brand), n_trials=25)

    best_params = study.best_params
    best_params["objective"] = "regression"
    best_params["metric"] = "mape"

    preds, acts, mape = walk_forward(df_brand, brand, best_params)

    results[brand] = {
        "best_params": best_params,
        "MAPE": mape,
        "preds": preds,
        "actuals": acts
    }

```

Gambar 3.25 Hyperparameter Optuna LightGBM

Pada gambar 3.25 adalah untuk mengevaluasi kinerja model pembelajaran mesin secara realistis juga mendorong pemilihan walk-forward. Model dapat mempelajari pola musiman, tren jangka panjang, dan perubahan pasar secara lebih akurat dengan memanfaatkan seluruh data dari setiap iterasi pelatihan. Walk-forward juga mencegah kehilangan data penting, yang sering terjadi pada rolling window yang membatasi jumlah data. Oleh karena itu, pendekatan maju menjadi yang lebih sesuai

untuk proses prediksi penjualan mobil. Metode ini sangat bergantung pada pola historis dan faktor eksternal yang terus berubah dari waktu ke waktu.

Untuk kebutuhan kompetisi LKTIN CARBON 07, teknik LightGBM digunakan. Fokus utama versi ini adalah feature engineering yang jauh lebih kaya dan kompleks. Selain interaksi tingkat lanjut seperti gabungan antara kurs, inflasi transportasi, dan Google Trends, fitur yang dibangun mencakup rolling mean, rolling standard deviation, lag multi-periode, dan transformasi musiman berbasis fungsi sinus-cosinus. Setelah itu, setiap fitur dinormalisasi menggunakan StandardScaler agar model lebih stabil terhadap variabel dengan skala yang berbeda. Pada target juga dilakukan transformasi log untuk menstabilkan distribusi data yang cenderung berubah-ubah. Metode rolling forecast berbasis titik per titik digunakan dalam pendekatan prediksi, sehingga model dilatih ulang pada setiap langkah prediksi. Konfigurasi model LightGBM yang sama digunakan oleh semua merek dalam versi ini karena LightGBM tidak menggunakan parameter penyesuaian.

Secara umum, perbedaan paling signifikan antara kedua kode terletak pada titik fokus masing-masing pendekatan. Kode yang digunakan dalam kompetisi sangat baik untuk algoritma seperti Random Forest dan XGBoost yang kuat terhadap kompleksitas fitur. Namun, karena parameter tidak dituning, performa LightGBM menjadi kurang optimal. Namun, pada kode peningkatan versi terbaru, kompleksitas fitur dikurangi dan digantikan dengan proses tuning yang intensif, sehingga LightGBM dapat mengeksplorasi kombinasi parameter terbaik dan memaksimalkan hasilnya. Perbandingan ini menunjukkan bahwa peningkatan kualitas model tidak hanya tergantung pada banyak fitur tetapi juga pada keseimbangan antara kompleksitas fitur, stabilitas data, dan parameter yang dioptimalkan dengan benar.

```

DAIHATSU:
  Best Params: {'n_estimators': 413, 'learning_rate': 0.1}
  MAPE (%) : 21.14
-----
HONDA:
  Best Params: {'n_estimators': 1068, 'learning_rate': 0.1}
  MAPE (%) : 36.04
-----
MITSUBISHI:
  Best Params: {'n_estimators': 876, 'learning_rate': 0.1}
  MAPE (%) : 24.07
-----
SUZUKI:
  Best Params: {'n_estimators': 1197, 'learning_rate': 0.1}
  MAPE (%) : 20.90
-----
TOYOTA:
  Best Params: {'n_estimators': 318, 'learning_rate': 0.1}
  MAPE (%) : 21.74
-----

```

Gambar 3.26 Evaluasi MAPE LightGBM

Pada gambar 3.26 menunjukkan perbedaan yang cukup signifikan dalam kinerja dari menggunakan parameter sebelumnya. Perbedaan ini disebabkan oleh perbedaan dalam strategi pemodelan. LightGBM menghasilkan nilai MAPE yang relatif seragam dan stabil untuk setiap merek pada kode peningkatan yang menggunakan Optuna. Nilai MAPE berkisar antara 20 dan 35%, dengan nilai terbaik untuk Suzuki (20.73%) dan Daihatsu (21.25%), dan nilai terendah untuk Honda (35.69%). Hasil ini menunjukkan bahwa proses tuning parameter berhasil memberikan konfigurasi model yang lebih sesuai dengan pola data setiap merek. Namun, performa LightGBM tidak sebaik yang diharapkan. Sebaliknya, pola kinerja yang sangat berbeda ditunjukkan oleh hasil kode versi kompetisi. Dalam versi ini, algoritma LightGBM berkinerja lebih buruk daripada dua algoritma lainnya, dengan MAPE berkisar antara 30 dan 52 persen, bergantung pada merek. Ini disebabkan oleh fakta bahwa algoritma ini tidak dituning sama sekali pada kode pesaingnya dan harus menangani banyak fitur yang rumit dan kompleks.

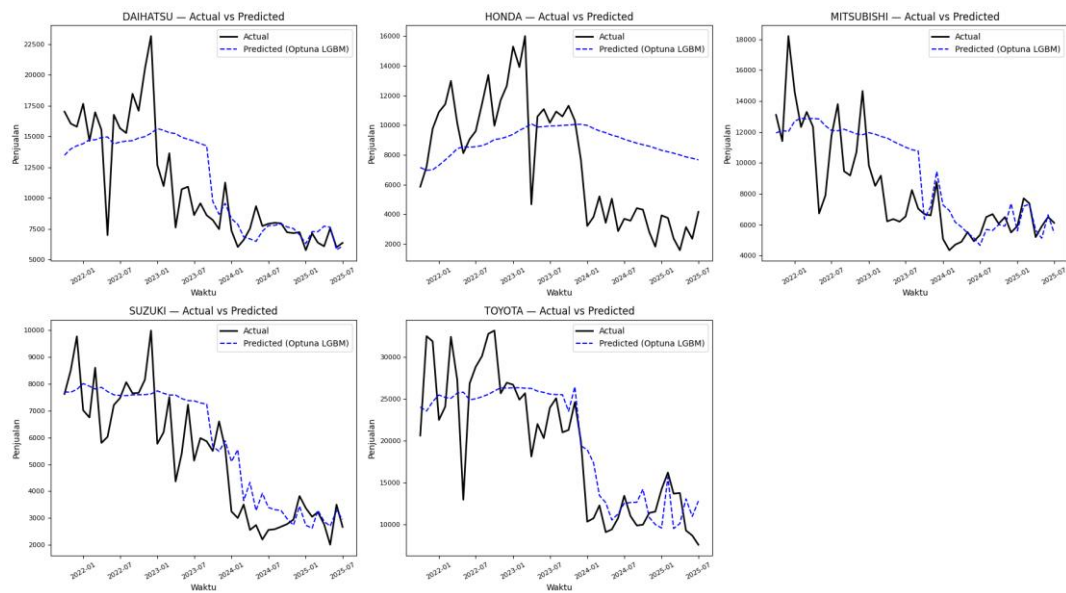
Hasil menunjukkan bahwa model peningkatan berbasis Optuna meningkatkan stabilitas performa LightGBM, sementara kode versi pesaing mendukung LightGBM kurang tanpa tuning parameter meskipun memberikan hasil terbaik untuk Random Forest dan XGBoost. Ditunjukkan oleh perbandingan ini bahwa keberhasilan sebuah

model sangat bergantung pada keseimbangan kualitas fitur, teknik praproses, dan kecocokan parameter terhadap karakteristik data.

Tabel 3.3 Perbandingan Evaluasi Hyperparameter LGBM

Metrik/Merek	MAPE (%) LightGBM		
	Optuna	Randomsearch	Hyperopt
Daihatsu	21.24	25.74	26.75
Honda	36.04	50.78	32.06
Mitsubishi	24.07	27.81	29.99
Suzuki	20.90	39.22	22.75
Toyota	21.74	23.27	38.83
Rata-rata	24.8	33.6	30.1

Pada Tabel 3.3 terlihat bahwa metode tuning hyperparameter menggunakan Optuna menghasilkan nilai MAPE rata-rata terendah untuk LightGBM, yaitu 24.8%. Nilai ini lebih baik dibandingkan Random Search yang memiliki rata-rata 33.6%, serta Hyperopt dengan rata-rata 30.1%. Pada level merek, Optuna juga memberikan kinerja yang konsisten lebih rendah pada hampir semua merek, misalnya Daihatsu 21.24% dan Suzuki 20.90%. Dengan demikian, pada konfigurasi model LightGBM dalam studi ini, metode tuning menggunakan Optuna dapat dianggap lebih optimal karena mampu menurunkan error secara signifikan dibandingkan dua metode lainnya.



Gambar 2.27 Actual vs Predicted (Optuna)

Pada gambar 3.27 perbedaan antara data penjualan mobil aktual dan hasil prediksi model Optuna LGBM untuk lima merek mobil DAIHATSU, HONDA, MITSUBISHI, SUZUKI, dan TOYOTA. Setiap grafik menunjukkan tren penjualan dari awal tahun 2022 hingga pertengahan tahun 2025. Waktu diwakili oleh sumbu horizontal, dan jumlah diwakili oleh sumbu vertikal. Garis biru putus-putus menunjukkan prediksi model, dan garis hitam solid menunjukkan data penjualan aktual. Secara keseluruhan, model Optuna LGBM dapat mengikuti pola penjualan dengan cukup baik, terutama untuk merek seperti HONDA dan TOYOTA, di mana prediksi sangat dekat dengan data yang sebenarnya. Untuk merek seperti DAIHATSU dan MITSUBISHI, ada beberapa perbedaan antara perkiraan dan kenyataan, terutama ketika terjadi kenaikan atau penurunan tajam, yang menunjukkan bahwa model mungkin kesulitan mengatasi perubahan ekstrim. Meskipun demikian, prediksi untuk SUZUKI menunjukkan stabilitas yang baik, meskipun tidak selalu akurat dalam mengidentifikasi perubahan cepat. Secara keseluruhan, grafik ini memberikan gambaran yang jelas tentang bagaimana model berfungsi untuk memproyeksikan penjualan mobil dan menunjukkan potensi model dan kelemahan di tengah-tengah dinamika pasar yang rumit.

Selain itu, dibandingkan dengan model LightGBM awal yang digunakan saat

kompetisi, hasil visualisasi menunjukkan peningkatan yang signifikan dalam akurasi. Nilai prediksi pada model sebelumnya sering kali tidak sesuai dengan data sebenarnya, terutama selama periode yang mengalami kenaikan atau penurunan tajam. Akibatnya, prediksi tidak dapat menunjukkan kondisi penjualan sebenarnya. Namun, setelah parameter dioptimalkan dengan Optuna, model baru menghasilkan prediksi yang jauh lebih dekat dengan nilai sebenarnya. Pada grafik, garis prediksi tampaknya mengikuti pola perubahan penjualan dengan lebih akurat dan konsisten. Dengan peningkatan ini, hasil prediksi menjadi lebih sesuai dengan keadaan sebenarnya karena proses penyesuaian parameter meningkatkan kemampuan model untuk mengidentifikasi perubahan dalam penjualan mobil.

Selain peningkatan kinerja, visualisasi gambar menunjukkan bahwa model Optuna LGBM mampu menggeneralisasi data baru dengan lebih baik. Hal ini ditunjukkan oleh kemampuan model untuk mengembangkan pola prediksi yang konsisten untuk berbagai merek dengan berbagai atribut penjualan. Model dari merek yang sangat beragam seperti MITSUBISHI dan DAIHATSU tetap dapat mengikuti tren utama meskipun terjadi fluktuasi besar, sementara merek dengan pola penjualan yang lebih stabil seperti HONDA dan TOYOTA dapat memproyeksikan nilai dengan sangat presisi. Perbaikan ini menunjukkan bahwa akurasi ditingkatkan melalui pengoptimalan parameter. Ini juga membuat model lebih responsif terhadap perbedaan perilaku pasar yang terjadi di masing-masing merek.

3.1.2 Kendala yang Ditemukan

Jumlah data yang terbatas merupakan kendala utama bagi tim peserta dalam menyusun kertas lengkap untuk kompetisi CARBON 07. Dataset yang digunakan hanya terdiri dari 53 baris, yang merupakan jumlah yang sangat kecil untuk keperluan analisis rangkaian waktu dan pembuatan model pembelajaran mesin. Oleh karena itu, ruang pembelajaran model terbatas, yang membuat model sulit untuk mengidentifikasi pola jangka panjang, variasi musiman, dan anomali pada data penjualan mobil.

Risiko underfitting dan overfitting meningkat jika jumlah datanya terbatas. Pada beberapa percobaan, model tidak dapat menangkap pola yang cukup kuat (underfitting), tetapi pada percobaan lain, model terlalu menyesuaikan dengan data latih yang sedikit. Kondisi ini berdampak langsung pada tingkat akurasi prediksi karena model dapat mengalami tingkat error yang tinggi ketika diuji pada data baru. Dengan masalah ini, proses pemilihan model dan pengaturan parameter menjadi lebih sulit. Ini karena setiap perubahan kecil dapat memengaruhi performa model secara signifikan.

Selain itu, telah ditunjukkan bahwa model LightGBM yang digunakan tim tidak cocok untuk konteks forecasting yang melibatkan dataset yang lebih kecil. Hasil pengujian menunjukkan bahwa nilai MAPE LightGBM jauh lebih besar daripada dua model lainnya XGBoost dan Random Forest yang justru memberikan kinerja yang lebih stabil. Ini menunjukkan bahwa untuk bekerja dengan baik, LightGBM membutuhkan jumlah data atau variasi fitur yang lebih besar. Salah satu masalah penting yang diangkat dalam analisis penelitian adalah kendala ini. Selain itu, ini membantu tim memahami keterbatasan metode pembelajaran mesin terhadap kondisi data tertentu.

Selain masalah tersebut, tim peserta harus menerapkan fair treatment pada ketiga model LightGBM, XGBoost, dan Random Forest dengan menggunakan parameter dasar yang sama pada tahap awal pemodelan. Metode ini digunakan untuk memastikan proses evaluasi antar model adil dan tidak bias terhadap algoritma tertentu. Namun, sementara XGBoost dan Random Forest masih mampu menghasilkan hasil yang cukup stabil, LightGBM justru menghasilkan akurasi terburuk di antara ketiganya. Ini karena LightGBM secara teknis lebih sensitif terhadap konfigurasi parameter dan biasanya membutuhkan tuning yang lebih spesifik. Ketidakcocokan ini semakin diperburuk oleh ukuran dataset yang kecil; model tidak memiliki cukup data untuk memaksimalkan kemampuan boosting-nya. Pada tahap evaluasi awal, nilai MAPE LightGBM jauh lebih tinggi

dibandingkan dua model lainnya, dan ini adalah salah satu alasan utamanya.

Kendala yang dihadapi selama kompetisi adalah pada saat sesi presentasi, di mana penyampaian materi terlalu berfokus pada aspek teknis. Kondisi ini membuat sebagian audiens dan juri yang tidak berada pada domain teknis tertentu menjadi kurang memahami konteks, alur, dan kontribusi penelitian. Akibatnya, pesan inti penelitian tidak tersampaikan secara maksimal dan kesan akademiknya menjadi terlalu berat.

3.1.3 Solusi atas Kendala yang Ditemukan

Tim peserta menemukan, sebagai solusi untuk penelitian mendatang, bahwa penggunaan dataset dengan jumlah baris yang lebih besar, idealnya 500 hingga 1.000 rows, sangat penting untuk mengurangi risiko overfitting maupun underfitting, terutama ketika menggunakan model berbasis boosting seperti LightGBM. Dataset dengan jumlah baris yang lebih besar akan memberikan variasi pola yang lebih kaya, yang memungkinkan model untuk mempelajari hubungan antar variabel dengan lebih stabil dan lebih akurat.

Selain itu, jumlah data yang memadai memungkinkan proses tuning parameter dilakukan secara optimal tanpa membuat model rentan terhadap fluktuasi kecil dalam data. Dengan penambahan data yang lebih banyak, model seperti LightGBM dapat menunjukkan kinerja terbaiknya, karena algoritma ini pada dasarnya dirancang untuk bekerja dengan baik pada dataset berukuran menengah hingga besar.

Oleh karena itu, perlu dilakukan upaya untuk memperluas cakupan data, baik melalui pengumpulan data dalam jangka waktu yang lebih lama, penggabungan lebih banyak variabel eksternal, maupun pencarian sumber data

tambahan yang relevan dengan dinamika penjualan mobil. Hal ini dapat meningkatkan kualitas hasil evaluasi model dan menghasilkan MAPE yang lebih akurat dari kondisi pasar yang sebenarnya.

Solusi dari kendala tersebut adalah dengan memahami karakteristik audiens dan menjelaskan penelitian dari sudut pandang yang lebih general terlebih dahulu, sebelum masuk ke bagian teknis. Pendekatan ini membantu membangun konteks agar semua pendengar dapat mengerti apa yang sedang dibahas. Selain itu, pembatasan istilah teknis, penggunaan analogi, visualisasi, serta penjelasan yang lebih naratif dapat meningkatkan keterlibatan audiens dan menjadikan presentasi lebih komunikatif.

3.2 Hasil Lomba/Kompetisi



Gambar 3.28 Menang Best Presentation

Pada gambar 3.28 adalah hari terakhir rangkaian acara di Universitas Jambi, tim peserta berhasil meraih penghargaan Presentasi Terbaik, seperti yang Salah satu dari dua kategori penghargaan ini diberikan kepada finalis yang menunjukkan kinerja terbaik dalam penyampaian hasil penelitian. Salah satu dari dua tim yang berhasil mendapatkan penghargaan ini adalah tim peserta yang terdiri dari tiga orang. Prestasi ini menunjukkan kemampuan tim untuk menyampaikan hasil penelitian secara terstruktur, jelas, dan meyakinkan kepada dewan juri.

Ketua Himpunan Mahasiswa Kimia Universitas Jambi (HIMKI) secara langsung memberikan penghargaan Best Presentation sebagai bentuk apresiasi atas presentasi ilmiah yang bagus. Momentum ini menandai kemampuan komunikasi akademik tim mereka dan keberhasilan mereka bersaing dengan finalis dari berbagai kota. Penghargaan ini meningkatkan pencapaian tim dalam kompetisi CARBON 07 dan menjadi pengalaman berharga yang menambah rekam jejak prestasi penelitian mereka.



Gambar 3.29 Sertifikat Carbon 7.0

Pada gambar 3.29 menunjukkan sertifikat resmi yang diterima oleh peserta Best Presentation pada kompetisi CARBON 07. Sertifikat ini menunjukkan kepada dewan juri bahwa presentasi hasil penelitian tim kami unggul. Hasil ini menunjukkan bahwa kelompok tidak hanya memiliki kemampuan untuk menyusun penelitian, tetapi juga memiliki kemampuan untuk menyampaikan hasil penelitian secara meyakinkan.